DOCUMENT RESUME

ED 178 969 CS 502 685

AUTHOR Brown, Kenneth L.; And Others
TITLE Assessment of Basic Speaking and Listening Skills: State of the Art and Recommendations for Instrument Development. Volume I.
INSTITUTION Massachusetts State Dept. of Education, Boston. Bureau of Research and Assessment.
PUB DATE Sep 79
NOTE 107p.; For related document, see CS 502 686

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS Basic Skills; *Listening Skills; Listening Tests; Mastery Tests; *Measurement Instruments; Minimum Competency Testing; Speech Communication; *Speech Evaluation; *Speech Skills; *State of the Art Reviews; Student Evaluation; Test Construction; Testing Programs; *Test Reviews

ABSTRACT

The first of two related documents, this volume reviews the state of the art of assessment in speaking and listening and recommends guidelines for evaluation and development of testing instruments. The first chapter recommends criteria for evaluating assessment instruments and practices; the second chapter reviews 71 published and unpublished instruments; the third chapter summarizes current testing practices throughout the country and compares them with the criteria listed in chapter one; the last chapter contains recommendations for the development of two kinds of instruments: a rating scale to assess speaking skills and a test to assess listening skills. (AEA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED178969

ASSESSMENT OF BASIC SPEAKING AND LISTENING SKILLS:

State of the Art

and

Recommendations for Instrument Development

Volume I

by

Assessment of Basic Speaking and Listening Skills

Project Staff

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Kenneth L. Brown, Project Director
University of Massachusetts, Amherst

Philip Backlund, Utica College, New York

Joanne Gurry, Marshfield High School, Massachusetts

Fred Jandt, SUNY Brockport, New York

A Report for

The Bureau of Research and Assessment

Massachusetts Department of Education

31 St. James Avenue

Boston, Massachusetts 02116

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Matthew H. Towle

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

September, 1979

CS502685

This project was developed with the assistance of the Bureau of Research and Assessment. However, the views expressed or materials included do not necessarily reflect Bureau or Department of Education Policy.

MASSACHUSETTS BOARD OF EDUCATION

Charles T. Grigsby, Chairperson
Ann H. McHugh, Vice-Chairperson

John Anthony
Mary Barros
John W. Bond
James L. Green
Howard A. Greis
Mary Ann Hardenbergh
Joseph C. Mello, Jr.
Edwin M. Rossman

Gregory R. Anrig
Commissioner of Education, Secretary

Ex Officio Member:

Edward C. McGuire, Chancellor
Board of Higher Education

BUREAU OF RESEARCH AND ASSESSMENT

Allan S. Hartman
Tracy Libros
Leslie S. May
Matthew H. Towle

The Massachusetts Department of Education insures equal employment/educational opportunities/affirmative action, regardless of race, color, creed, national origin or sex, in compliance with Title IX. Publication #11527 approved by Alfred C. Holland, State Purchasing Agent.

TABLE OF CONTENTS

VOLUME I

INTRODUCTION i

Chapter

I. CRITERIA FOR EVALUATING INSTRUMENTS AND PROCEDURES THAT ASSESS SPEAKING AND LISTENING SKILLS 1

II. STATE OF THE ART: SPEAKING AND LISTENING INSTRUMENTS 9

Previous Related Research 9

Review of Instruments 13

Summary 49

III. STATE OF THE ART: PRACTICES IN OTHER STATES AND THEIR JURISDICTIONS 50

Review of Practices 50

Summary 64

Comparison with Criteria 66

IV. RECOMMENDATIONS FOR THE DEVELOPMENT OF INSTRUMENTS AND PROCEDURES 68

Assessing Speaking Skills 68

Assessing Listening Skills 73

REFERENCES 75

VOLUME II

Appendixes

A. INSTRUMENT ABSTRACTS AND RATING FORMS 86

B. 1978-1979 SURVEY OF THE STATUS OF SPEECH COMMUNICATION EDUCATION IN MASSACHUSETTS PUBLIC SECONDARY SCHOOLS 462

C. SELECTED BIBLIOGRAPHY ON SPEECH COMMUNICATION IN THE ELEMENTARY SCHOOL 477

D. SELECTED BIBLIOGRAPHY ON SPEECH COMMUNICATION IN THE SECONDARY SCHOOL 484

LIST OF TABLES

1. Speaking Instrument Ratings for Criteria 1 through 9 and 11 through 15 29

2. Listening Instrument Ratings for Criteria 1 through 9 and 11 through 15 31

3. Speaking and Listening Instrument Ratings for Criteria 1 through 9 and 11 through 15 33

4. Skills Stressed in Speaking Instruments (Criterion 10) 35

5. Skills Stressed in Listening Instruments (Criterion 10) 37

6. Skills Stressed in Speaking and Listening Instruments (Criterion 10) 38

7. Rank Order of Speaking Skills 44

8. Rank Order of Listening Skills 45

9. Potential Model Instruments that Assess Fifty Percent or More of the Speaking and Listening Skills by Category and Level 47

10. Skill Areas Assessed in State Minimal Competency Testing Programs 55

LIST OF FIGURES

1. Instrument Abstract Form 15
2. Instrument Rating Scale 20
3. Instruments Reviewed by Category and Level 23
4. Instrument Types 27

INTRODUCTION

On January 23, 1979, the State Board of Education approved a document, "Regulations for Implementation of the Basic Skills Improvement Policy."[1] In addition to specifying skills in reading, writing, and mathematics, the regulations specify eleven listening and fourteen speaking skills that must be included in basic skills improvement programs throughout the Commonwealth. These skills are as follows:

Listening
- (a) Basic Listening Skills
  1. Recognize words and phrases used by the speaker
  2. Indicate why the speaker can or cannot be understood
- (b) Understanding What You Hear
  1. Understand spoken words and ideas
  2. Identify and understand main ideas
  3. Associate important details with main ideas
  4. Understand descriptions of events and experiences
  5. Understand speaker's purpose
- (c) Using What You Hear
  1. Understand and respond to survival words used in emergency situations
  2. Summarize information and draw conclusions
  3. Recognize when words and phrases are used to convince or persuade
  4. Follow straightforward directions

Speaking
- (a) Basic Oral Communication Skills
  1. Use words and phrases appropriate to the situation
  2. Speak loudly enough to be heard by a listener or group of listeners
  3. Speak at a rate the listener can understand
  4. Say words distinctly
- (b) Planning, Developing and Stating Spoken Messages
  1. Use words in an order that clearly expresses the thought
  2. Organize main ideas for presentation
  3. State main ideas clearly
  4. Support main ideas with important details
- (c) Common Uses of Spoken Messages
  1. Use survival words to cope with emergency situations
  2. Speak so listener understands purpose
  3. Ask for and give straightforward information
  4. Describe objects, events and experiences
  5. Question others' viewpoints

---

[1] Since it covers more information than bears directly on our focus on assessment of speaking and listening skills, the complete document is not included in this report. Copies are available from The Bureau of Research and Assessment, Department of Education, Commonwealth of Massachusetts, 31 St. James Avenue, Boston, MA 02116.

Through the regulations, the Board requires that student achievement of the speaking and listening skills must be evaluated no later than the end of the 1981-82 school year.

Each school district can establish its own evaluation plan at the elementary school level. Secondary schools have the option of using one or more of the following evaluation instruments: (a) instruments available from the Department of Education; (b) commercially available instruments approved by the Department of Education; (c) locally developed evaluation instruments approved by the Department of Education as being comparable to either (a) or (b). The regulations also specify that secondary level students shall first be evaluated no later than the beginning of grade nine.

To begin implementing these regulations concerned with speaking and listening, members of the Department of Education's Bureau of Research and Assessment, in cooperation with members of an advisory Speaking/Listening Committee, invited the project director to assist with the Policy's evaluation component. The project director recruited three professional colleagues -- all knowledgeable about basic speaking and listening skills and about assessment -- to form the project staff.

Purposes

The task assigned to the project staff was to prepare a document that reviews the state of the art of assessment in speaking and listening and that recommends guidelines for evaluating current instruments and developing new ones.

We interpreted this charge to consist of four basic objectives: (a) to recommend criteria for evaluating instruments and procedures that assess speaking and listening skills; (b) to review existing instruments; (c) to review assessment practices in other states and their school districts; (d) to recommend specifications for the Department of Education instruments for speaking and listening.

We defined instruments as published and unpublished tests, inventories, checklists, rating scales, observation schedules, situational tasks, and speaking samples. This definition includes, but is not confined to, standardized tests. The term procedures was defined to mean methods of administering instruments (e.g., one-to-one, small group, entire class), responding to instruments (e.g., paper-pencil, oral-aural, pointing to objects), and scoring responses (e.g., analytic, holistic, etc.).

To summarize our purposes in question form, we are asking:

- What criteria should be used to evaluate instruments and procedures?
- What instruments and procedures are currently available?
- How do these instruments and procedures compare with our criteria? What strengths, weaknesses, gaps appear?

- What existing instruments can be recommended to the Bureau of Research and Assessment for possible adoption?
- How are speaking and listening skills evaluated in other states and their school districts?
- What kinds of instruments and procedures need to be developed? What is the feasibility of developing such instruments?

Procedures

The project staff accomplished its goals through a series of activities that began in March, 1979. First, the staff met in Amherst to clarify the task, organize the review, and develop criteria for evaluating assessment efforts. Development of the criteria was the principal agenda item of this first two day meeting.

Next, the staff members worked independently on predetermined assignments: reviewing assessment research; revising the criteria; locating, ordering, abstracting, and analyzing instruments; conferring with supervisors of basic skills programs in other states and school districts and with other professionals; writing periodic progress reports; exchanging information and evaluating data. In late May, the project director met with members of the Bureau of Research and Assessment and with members of the Speaking/Listening Subcommittee of the Communication Task Force to present a progress report.

In early June, the project staff met for three days to synthesize research and prepare final recommendations. In late June the project director met with members of the Bureau and with the co-chairpersons of the advisory Speaking/Listening Committee to report the recommendations and to consult on development of a "Request for Proposals" for dissemination to potential contractors.

In Chapter I we recommend criteria for evaluating assessment instruments and practices. Chapter II presents our review of existing instruments in light of these criteria. Chapter III reviews assessment practices in other states and compares those practices with our criteria. Chapter IV contains our final recommendations for instrument development.

Appendix A includes abstracts and ratings of each of the seventy-one instruments we reviewed. These materials comprise the "raw" data for the analysis contained in Chapter II. Appendix B presents a survey of speech programs in the public secondary schools of the Commonwealth. This survey was conducted in the 1979 spring term as an independent investigation, but we include it here because it contains information relevant to instrument development. Appendixes C and D are selected bibliographies of sources on teaching speaking and listening skills. We hope they will be useful to those who seek continuing education in these important basic skills areas.

Readers familiar with the document "Regulations for Implementation of the Basic Skills Improvement Policy" will note in this report that the project staff renumbered the speaking and listening skills in Chapters II, III, and IV. We found it simpler to discuss the skills with each set numbered consecutively. Figure 1 reflects the renumbering. Chapter I is an exception to this practice; there we refer to the specific skills less frequently, so we identify them the same as in the "Regulations."

We hope that the perspective provided by this report will stimulate further research in and assessment of speaking and listening skills.

Acknowledgements

Several individuals deserve recognition because of their contributions to the project. Special thanks go to Tracy Libros of the Bureau of Research and Assessment. She was instrumental in clarifying our task, securing instruments, and advising us at all stages. She gave us great latitude and never imposed her own views on us.

We are grateful to Virginia Ross of Marshfield High School, Eileen Malloy of Whitman-Hanson Regional High School, and Allan Hartman of the Bureau of Research and Assessment whose advice helped shape our report.

Finally, we wish to warmly acknowledge the test publishers and state and local supervisors who provided us with materials on which this report is based.

Kenneth L. Brown
Summer 1979

CHAPTER I

CRITERIA FOR EVALUATING INSTRUMENTS AND PROCEDURES THAT ASSESS SPEAKING AND LISTENING SKILLS

In this section we present criteria that are important in evaluating assessment instruments and procedures for listening and speaking. These criteria can be applied to published and unpublished instruments and procedures; to practices in other states and school systems that are already evaluating students and curricula; to future instruments and procedures that might be developed; and to program plans submitted by local school districts for approval by the Department of Education.

Before presenting the criteria, we want to make two prefatory statements regarding our conceptual view. The first concerns relationships among listening, speaking, reading and writing skills. The second concerns the role of prediction in assessing communicative abilities.

Relationships Among Listening, Speaking, Reading and Writing Skills[1]

As we review the state approved basic skills in reading, writing, speaking and listening, we note in the published lists a strong relationship among these skills, particularly between expressive communication skills (speaking and writing) on the one hand and between receptive communication skills (listening and reading) on the other hand. Examples of commonalities in expressive skills include: supporting ideas with important or relevant details (speaking skill #b-4 and writing skill #a-2); having and making clear a purpose (speaking c-2; writing b); organizing ideas (speaking b-2; writing c-2); and demonstrating awareness of the intended audience (speaking a-1 through 4; writing d). In fact ten speaking skills correspond closely with nine writing skills.

Examples of commonalities in receptive skills include: recognizing and identifying meanings of words and phrases (listening skill #a-1 and reading skill #a-1 and a-2); identifying and understanding main ideas (listening b-2; reading b-3); associating important details with main ideas (listening b-3; reading b-3); determining the communication purpose (listening b-5 and c-3; reading d-2), and following directions (listening c-4; reading b-2). In all, eight listening skills correspond closely with seven reading skills.

---

[1] The specific skills referred to in this section are found in "Regulations for Implementation of the Basic Skills Improvement Policy," Commonwealth of Massachusetts Department of Education, January 23, 1979, pp. 3-4. We do not renumber the skills here as we do elsewhere in this report.

Such correspondence is noteworthy when one considers that the Communication Task Force which drafted the list of skills was comprised of three subcommittees -- Reading, Writing, and Speaking/Listening. Each subcommittee defined the skills relatively independently. This correspondence confirms the Task Force's basic position that communication competency is a conceptual entirety and that all the skill areas of listening, speaking, reading and writing are essential in determining communication competency. This correspondence also demonstrates that teachers and administrators agree to define expressive and receptive skills as interrelated.

This agreement among members of the Communication Task Force strongly suggests a high degree of interrelatedness among the various communication skills. Oral language development serves as a base for developing skills in reading and writing. Reciprocally, the ability to comprehend written material through reading and to express oneself through writing contribute to maturity in speaking and listening.

But this relationship must be treated with caution. It is tempting to assume that skills learned through one mode of communication, e.g., writing, can be transferred easily to another mode, e.g., speaking. An extension of this conclusion is that assessment instruments and procedures suitable for writing might also be suitable for speaking. One might even go so far as to suggest that assessment of reading and writing skills is all that is necessary to assure achievement of communication competency.

Such conclusions (and we have heard these stated) are faulty. The best tests are those that assess behavior directly. Therefore, a reading or writing test may assess and be valid for reading or writing competencies, but to make a judgment of speaking and listening competencies from the same test would be highly and dangerously inferential.

Assessment of communication skills must account for significant differences between oral and written language and hence for skills that are unique to speaking and listening and to reading and writing. We see at least three important kinds of differences that must be accounted for in the assessment process. We characterize these distinctive features of oral communication as time, medium, and relationship.

Time. The oral communication situation is characterized by a sense of immediacy. Time becomes a source of pressure. The speaker is pressed to express ideas in language as he/she directly confronts the audience. Time to look for the right words or phrases is limited (hence, the higher incidence of grammatical errors in oral communication). Re-stating, repeating and summarizing information are more necessary. The rate of communication is controlled more by the speaker than by the listener. Referability is limited. Unless audio and video recording devices are employed the speech record is impermanent; whereas in writing the writer has the luxury of carefully thinking out words and phrases and editing them until they are satisfactory before submitting them to the receiver. Even then, response is not immediate and does not require the give-and-take of oral communication.

Medium. While the writer employs graphic symbols, the speaker relies on sound and nonverbal symbols. This creates the need for different skills. The writer must demonstrate skills in capitalizing, punctuating, spelling and writing legibly to convey a message that is intelligible. The writer uses other contrasts between visible shapes and blank spaces such as color, intensity, different letter size, underlining. On the other hand the speaker employs voice quality (harsh-smooth), volume (loud-soft), pitch (high-low), rate (fast-slow), intonation (rising-falling), stress (emphasis), juncture (pause), and bodily actions to communicate messages to the listener. The different media require unique skills.

Relationship. The speaker-listener encounter is a face-to-face relationship that is characterized by directness. When addressing listeners the speaker's style of language is unique -- more concrete than abstract, more personal, more cognizant of time, place and occasion. Vocabulary tends to be simpler, and the density of ideas is greater. The listener will hear the speaker use more self reference terms (first person pronouns), more pseudo-quantifying terms (if, however, except), and more terms indicative of consciousness of projection (apparently, seem, appears). The language employed by the speaker, then, differs from the language employed by the writer in amount of abstraction, difficulty of comprehension, and psychogrammatical features (Blankenship, 1968, 1974; DeVito, 1965, 1966, 1967).

We note these distinctive features of oral communication because they emphasize the need for direct teaching and assessment of speaking and listening skills. We do not automatically teach speaking when we teach writing. We do not automatically assess achievement of speaking skills when we assess achievement of writing skills.

Reading, writing, listening and speaking skills are indeed interrelated. The competent communicator will demonstrate skills in all these areas. But instructional programs and assessment procedures must account for important distinctive skills as well as interrelated skills. The results of assessing reading and writing skills should not be used to indicate achievement of speaking and listening skills.

Predictability

Our second prefatory statement is a caveat that pertains to all testing. It concerns the predictive character of assessment. Determining competency in oral communication is complex because it is dependent on the interaction of different factors including: the speaker's and listener's purpose or task in communication; the topic or subject being talked about; the attitudes, experiences, maturity, skills and knowledge background of both the speaker and listener; the time, place and preceding events of the communication setting. These and other factors converge in more specific communication situations such as the following: the student explaining job requirements to a new employee at work; the student describing an accident to a policeman; the student obtaining information from public officials about local laws and regulations; the student answering a doctor's questions about an illness he/she has; and the student describing the viewpoint of a friend with whom he/she disagrees. Each student will handle these situations with different degrees of success. Success will depend on the student's knowledge of and skill in handling the purpose, topic of talk, the person(s) involved and the time and place. To determine definitively whether or not a person has achieved a given level of competence would involve following the student around and observing his/her performance in all of these kinds of situations.

Judging achievement of competency on the basis of predetermined speaking and listening skills makes the task somewhat easier because these skills force us to focus on relatively specific kinds of behaviors. Given our list of skills, assessors need not be concerned with skills such as expressing feelings or performing various communication rituals. Assessors will need to focus on such communication purposes as giving explanations, directions and instructions. But even here the specific situations for accomplishing these purposes are numerous. For example, a student may need to give directions in a wide range of situations including telling a new fellow employee how to complete a given work task; telling a friend or stranger how to reach a given destination; telling a younger brother or sister how to play a game; telling a friend how to repair a piece of equipment; telling a younger child what to do in case of a fire or another emergency. Even with the purpose of communication narrowed, all the possible situations in which the student might be expected to demonstrate competence are too numerous to be assessed one-by-one.

Thus we cannot say that any single assessment instrument or procedure will give us a definitive certification of a student's level of competency in speaking and listening skills. Faced with the limitations imposed by the situational character of communication behavior and by practical factors (such as time and money allotted for testing), we can only infer a given student's ability to communicate competently in situations beyond the assessment. This is both a strength and a weakness of assessment. It is a strength in that it fits well with the overall goal of competency education -- that we want to educate students who, we assume, will function at least in some minimally competent manner outside as well as inside the classroom. It is a weakness in that one can never know for certain how competent a student is, and no testing procedure devised will give a definitive answer. We recognize this as a common problem of assessment, but we believe it is important to recognize that assessment ultimately looks beyond demonstration of skills attainment in one setting to a reasonable expectation that skills will be utilized in future similar settings.

Criteria[1]

Our criteria are organized around content specifications and technical specifications. Content considerations deal primarily with the substance, the basic nature, of speaking and listening instruments and procedures. The technical considerations deal with such matters as reliability, validity and information on administration. All the criteria are important, but the most critical are starred (*). Some criteria are rather self-evident and are presented briefly. Other criteria require elaboration.

Content Specifications

*1. Stimulus materials should require students to demonstrate skill as a speaker or listener. Our foremost concern is that assessment instruments and procedures should involve students in speaking and listening acts. This may sound obvious, but we are familiar with evaluation measures in oral communication which fail to assess speaking and listening skills. Some measures provide attitudinal, disclosure or style correlates; some measure apprehension or anxiety in communication situations; and some test the student's knowledge of communication principles rather than skill in communicating. These measures frequently take the form of paper and pencil tests. We have no quarrel with such measures. They provide much useful information, but they do not necessarily assess skill in speaking and listening. Such measures may be used in judicious combination with others that do engage students in speaking and listening acts, but with this criterion we are excluding instruments and procedures that are totally void of oral or aural performance.

*2. Assessment instruments and procedures should clearly distinguish speaking and listening performance from reading and writing ability. By this, we mean (a) a speaking/listening test should assess just that; it should not make inferences from reading/writing assessment; and (b) the directions and responses for a speaking/listening instrument should not be mediated through reading/writing modes.

---

[1] These criteria were developed by the Project Staff for this project. They were also subsequently approved and endorsed, in abbreviated form, by the Educational Policies Board and the Administrative Committee of the Speech Communication Association.

*3. Assessment instruments and procedures should be free of sexual, cultural, racial, and ethnic content and/or stereotyping. In this criterion we are emphasizing section 40.05:(6) of the "Regulations for Implementation of the Basic Skills Improvement Policy (January 23, 1979)." Instruments and procedures that are utilized should not discriminate on the basis of race, sex, religion or national origin.

*4. Assessment should confirm the presence or absence of skills, not diagnose reasons why students demonstrate or fail to demonstrate the skills. This criterion indicates that the primary purpose of basic skills assessment is to ascertain whether or not students have achieved the given skills. Some existing measures are diagnostic of student problems. They attempt to test for personality, social adjustment, intelligence or anxiety factors that may be related to a given behavior. But our assessment is not intended to be diagnostic of student problems. When students fail to demonstrate given skills, diagnostic instruments might be employed to identify sources of problems.

5. Assessment should emphasize application of speaking and listening skills in familiar situations. This criterion implies three features. First, stimulus materials should contain communication situations that most students will recognize easily. Explaining to a friend how to make a simple product is likely to be more familiar than criticizing the mayor's plan for urban renewal.

Second, the responses students give should be evaluated more for demonstration of speaking and listening skills than for accuracy of information or the content of the response. For example, if a student explains to a friend how to construct a simple concession stand for the school fair, his/her demonstration of communication skills (clarity and distinctness of speech, choice of words, organization of ideas) should be rated, not the accuracy of his/her knowledge of construction tools, materials and methods.

The third aspect of this criterion is closely related to the two previous factors. The stimulus material must provide a familiar hypothetical setting or circumstance through which the student demonstrates his or her ability to apply skills. But the stimulus should not confine the student's responses to the most narrow range of his or her experience. For example, by virtue of knowledge or experience, some students might be very adept at explaining aspects of their favorite sport, craft, or hobby, but be unaware of how to generalize information-giving or direction-giving to situations beyond their own content-specific boundaries. Stimulus materials, therefore, should not only facilitate demonstration of oral and aural skills in situations, but also should allow the student to use and apply speaking and listening skills in a range of common communication experiences.

6. Assessment should allow for optional communication settings rather than limit students to one setting. Oral communication settings can be classified generally as interpersonal, small group, public, and mass. This system of classification is based on factors such as the number of participants present (ranging from one-to-one through one-to-many), the kind of interaction among participants (ranging from direct through indirect), and the style of discourse (ranging from spontaneous through formal). Conversations are a typical form of interpersonal communication. Group discussions to solve problems or plan courses of action are examples of small group communication that students would be familiar with. Public communication consists of one person addressing a group of listeners or an audience. Mass communication involves listening to or giving messages that are mediated by radio, television and film.

The basic speaking and listening skills approved by the Board of Education do not refer to particular settings. As defined, the skills are free of a communication context. But assessment will involve evaluating the presence or absence of skills in one or more settings. We believe assessment should not limit the student to one particular setting. Whatever instruments or procedures are employed, they should not require students to demonstrate the skills only in public speaking settings, only in small group discussion settings, or only in informal conversations or interviews.

7. Assessment should involve situations that allow for a range of responses. The testing procedure should recognize that a variety of responses can be considered appropriate in everyday communicative situations. Therefore it should allow for a range of acceptable responses.

Technical Specifications

8. Assessment should demonstrate that outcomes are more than just chance evidence, i.e., that the same results would be achieved if the assessment process were repeated. Assessment should allow for inter-observer agreement with at least two raters. In short, assessment should be reliable.

9. Assessment should provide results that are consistent with other evidence that might be available, e.g., the teacher's experiences with the students. This is essentially a matter of concurrent validity.

*10. Assessment should correspond as closely as possible with the given speaking and listening skills identified in the Basic Skills Improvement Policy. This is essentially a matter of content validity.

11. Assessment procedures should be standardized and detailed enough so that individual responses will not be dependent on an

administrator's skills in administering the procedures. Students should not fail because the stimulus materials are administered poorly. The instrument should be one that can be administered by classroom teachers with relatively modest training in administration, use and interpretation of the instrument.

12. Assessment procedures should be as free as possible of such negative outcomes as undue stress or anxiety. We recognize that the evaluation process itself produces anxiety for some students and that evaluation of oral communication is particularly troublesome to students. No speaking or listening assessment is likely to be free of stress or tension for the student. But assessors, and the instruments and procedures they employ, should approximate the recognized stress level of oral communication; they should not increase or eliminate it.

13. Assessment procedures should be practical in terms of cost and time. This is a matter of cost effectiveness and administrative feasibility.

14. Assessment should involve simple equipment such as most school districts have on hand. This should eliminate expensive and elaborate laboratory equipment and materials.

15. Assessment should be suitable for the target population of the Basic Skills Improvement Program. It should be aimed at the child who is developing along "normal" lines. It should be suitable for students between seventh and twelfth grades, and it should be suitable for today's students.

CHAPTER II

STATE OF THE ART OF SPEAKING AND LISTENING INSTRUMENTS

Our purposes in this chapter are to briefly review previous research on the assessment of speaking and listening and to identify strengths and weaknesses of existing instruments.

PREVIOUS RELATED RESEARCH

The assessment of speaking and listening behavior has been addressed in at least four bodies of literature. The first pertains to speech rating scales, the second to functional communication measures, the third to listening tests, and the fourth to oral language measures for bilingual students. We will examine the research under these headings.

Speech Rating Scales

Investigators have studied the rating of speech behavior in basic speech classes (Becker, 1962; Becker and Cronkite, 1965; Brooks, 1957; Bryan and Wilkie, 1941; Clevenger, 1964; Cronen and Eckles, 1975; Fotheringham, 1956; Heinberg, 1963; Knower, 1929; Monroe, Remmers, and Lyle, 1936; Norrelle, 1934; Price, 1964; Stevens, 1928; and Thompson, 1943, 1944) and in competitive speech events (Ellis and Minter, 1967; Giffin, 1959; Hufford, 1965; Knower, 1940; Krueger, 1949; Kruger, 1955; McCroskey and Camp, 1964, 1966; Scott, 1955; Williams, Clark, and Wood, 1966). Dating back to 1928, these investigations reflect concern for developing speech rating scales to evaluate public speaking performance. Scale construction, scale independence, and rater reliability were three important issues addressed.

Early studies were preoccupied with scale construction -- the identification and definition of traits to be evaluated, the number of scale steps, and the relative importance of items. Some of the early scales contained as many as sixteen traits. Developers of these scales assumed these traits represented different and discrete aspects of speaking ability. Thus, raters could measure discrete components of speaking effectiveness.

Later studies challenged this assumption. By using a factor analytic research procedure, Becker (1962) found that a scale containing eleven items really only discriminated three factors: content-analysis, delivery, and language. He concluded that raters make a limited number of judgments regarding a few general aspects of performance.

Price (1964) and Cronen and Eckles (1975) found similar results. After compiling a list of thirty-five items, Price found teachers made judgments on the basis of six factors: content, clarity of utterance, bodily action, personality, language, and voice. Cronen employed a list of seventeen items to find they discriminated among three factors: reasoning, delivery, and organization.

Studies have also focused on the reliability of ratings. Generally, investigators have attributed low rater reliability to differences in raters' predispositions (Waldkoetter, Urrey, Martinson, 1965), personal biases (Henrikson, 1940), and anchors of judgment (Madden and Bourdon, 1964). But other factors affect reliability such as the number of judges (Cronen and Eckles, 1975; Knower, 1940; Price, 1964), inadequate use of all the steps in a rating scale (Becker and Cronkite, 1965), the number of traits on a scale (Clevenger, 1964), and the training of judges (Clevenger, 1964). These factors interact in a complex way to influence reliability; but the research also indicates some direction for increasing reliability. First, consider the type of speaking to be rated. It appears to be easier to achieve reliability for original and extemporaneous speaking than for more artistic types of performance (Knower, 1940). Second, use the full range of steps on a rating scale, but note that there is little increase in reliability above a six step scale and little decrease below a four step scale (Becker and Cronkite, 1965). Third, rate more than one trait such as general effectiveness (Clevenger, 1964). Fourth, use trained raters. This last principle requires elaboration.

The training of raters can be defined in two ways: by similarity of background and by training in use of the rating scale. The literature cited above has defined training the first way. Raters have been considered trained by virtue of their depth of study and amount of teaching experience in the field of speech. The studies cited above have not isolated the effect of training in the use of the rating scale on reliability. Studies of the evaluation of writing, particularly of holistic scoring methods (Cooper, 1977; Diederich, 1974; Follman and Anderson, 1967; and Moslemi, 1975) indicate that reliabilities in the high eighties and low nineties can be achieved when raters are from similar backgrounds and are trained in the particular method of evaluation. Therefore, one key in achieving satisfactory reliability is to train raters to interpret a given scale and to practice scoring on a sample population of students.

Another key is to employ more than one judgment. Research in writing (Cooper, 1977; Diederich, 1974) indicates this can be achieved by requiring two pieces of a student's writing and two independent ratings of each piece. Research in speech (Cronen and Eckles, 1975; Price, 1964) indicates that for one performance, multiple judges are necessary to achieve higher reliabilities.

Functional Communication Measures

A second body of literature comes from speech and related disciplines. This body is concerned with teaching and assessing functional communication (Allen and Brown, 1976; Allen and Wood, 1978; Bassett, Whittington, Staton-Spicer, 1978; Larson, 1978; Larson, Backlund, Redmond, Barbour, 1978; Wiemann, 1978; Wilkinson, 1965, 1971; Wood, 1977a and b). Functional communication involves the ability to demonstrate knowledge of socially appropriate communication behavior and the ability to achieve particular outcomes in a variety of communication situations. While early efforts to assess speaking ability focused on public speaking performance, investigators concerned with functional communication have focused on speaking and listening performance in an array of situations we encounter daily -- conversing with friends, solving problems in small groups, interviewing, giving instructions to peers, persuading others, etc. (Larson, Backlund, Redmond and Barbour (1978) discuss six common dimensions of communicating in different situations: language development, interpersonal competence, communication anxiety, nonverbal behavior, empathic ability, and listening. Larson et al. (1978) also classify and review ninety measures that tap one or more of these dimensions. Grouped by function and age level, the measures include a variety of tests, observation systems, and assessment procedures. Our review includes those measures that correspond with our speaking and listening skills.

Listening

The third body of literature focuses on listening. Listening research developed most rapidly in the fifties and sixties. It covers numerous aspects of the listening process including definition, underlying dimensions, methods of teaching and improving listening ability, and methods of assessment (Duker, 1971a and b). Duker's 1968 Listening Bibliography lists 1,332 references. While this flurry of activity decreased in the seventies, teaching and assessing listening still concern educators (Lundsteen, 1979; Wilkinson, 1976). This concern stems from a continuing study of relationships between reading and listening (Devine, 1968, 1978; Lundsteen, 1979; Sticht, 1979; Sticht et al., 1974) and leads to testing measures similar to those in reading.

According to Duker (1971b) the first test of listening skills was prepared by Donald D. Durrell and Helen B. Sullivan of Boston University in the 1930's. Publishers doubted that a listening test would be marketable, so the test was named the Durrell-Sullivan Reading Capacity Test. It has been used since its publication in 1937. Many more listening tests have appeared since then, but we do not know how many of these tests exist. Lundsteen (1979) reviews eight standardized tests, five unpublished tests and twelve tests that accompany published listening materials. Duker (1968) annotates sixty-five entries on unpublished tests. Our review of twenty-one listening measures and sixteen measures that assess both speaking and listening shows that listening assessment is frequently included in instruments that assess other skills.

Some listening tests have stimulated forceful, and sometimes shrill, criticisms (Carroll, 1971; Kelly, 1971). Backlund (Larson et al., 1978) sees these criticisms revolving around definitions of listening, validity, and problems compounded by measuring an essentially covert act. But responses to criticisms have also been forceful (Brown, 1971; Lundsteen, 1979). Current assessment procedures can benefit from these different views (Lundsteen, 1979; Wilkinson, 1965; Wilkinson et al., 1976).

Oral Language Measures for Bilingual Students

Research in the teaching of English to bilingual students has produced a fourth body of literature on assessment of speaking and listening (Aitken, 1977; Burt and Dulay, 1978; Chastain, 1979; De Avila and Duncan, 1977; Duncan and De Avila, 1978; Eslava and Lawson, 1979; Fairfield Board of Education, 1968; Findley, 1977; Fischer, 1978; Lowe, 1976; Matluck and Mace-Matluck, 1975; Northwest Regional Educational Laboratory, 1976; Sousa, 1971, vols. 1 and 4; Snow and Perkins, 1979). This literature reflects a concern with measuring language dominance, language proficiency, and language as a communication tool. Language dominance refers to the primary language in which a student interacts most often in the home. Language proficiency is defined as the degree to which an individual demonstrates linguistic competence -- i.e., knowledge of phonology, grammar, vocabulary, and semantics -- regardless of how the language may have been achieved. Like functional communication, language as a tool involves ability to use language for particular ends.

Cloze tests have been advocated as one way to measure overall proficiency with language (Aitken, 1977). Other interesting methods call for rating students on their performance of relevant tasks such as telling or writing stories, asking questions, interviewing, conversing about job related topics, creating dialogues, and role-playing (Burt and Dulay, 1978; De Avila and Duncan, 1977; Duncan and De Avila, 1978; Eslava and Lawson, 1979; Findley, 1977; Lowe, 1976; Matluck and Mace-Matluck, 1975; Snow and Perkins, 1979). Most of these measures are criterion referenced. Some are included in our review.

REVIEW OF INSTRUMENTS

Purposes

We reviewed instruments with two purposes in mind. First, we wanted to determine the degree to which existing instruments met our criteria and therefore were suitable for adoption. Frankly, we were skeptical at the outset that suitable instruments existed. Our own knowledge of testing in speaking and listening, advice we gathered from colleagues, and published reviews of tests led us to be cautious, especially with regard to assessing speech. For example, Loban (1976) indicated "no published test attempts to appraise the spoken word, although a few do seek to evaluate listening" (p. 45). At the same time we knew of recent efforts to assess both speaking and listening. We believed that by expanding our review beyond published tests we might find a suitable measure.

Our second purpose was to identify instruments that contain features which could be incorporated in the development of new measures. Hence, if a given instrument failed to meet the criteria we established earlier, we asked: To what extent does the instrument approach our criteria? Are some features stronger than others? Can some features be adapted to our purposes? How can this instrument contribute to the development of a more suitable measure?

Scope

With these two purposes in mind, we began to gather various instruments. We first consulted reviews of published and unpublished instruments (Bloom, Hastings, and Madaus, 1971; Bonjean, Hill, McLemore, 1967; Buros, 1978; Duker, 1971a and b; Fagan, Cooper and Jensen, 1975; Grommon, 1976a; Johnson vols. 1 and 2, 1976; Johnson and Bommarito, 1971; Larson, Backlund, Redmond, Barbour, 1978; Lundsteen, 1979; Simon and Boyer, 1974; Webb, Campbell, Schwartz and Sechrest, 1965; Wilkinson, 1965). We also located instruments by reviewing Human Resources Research Organization (HumRRO) literature, dissertation abstracts (see especially Black, 1977; Granowsky, 1976; Grossman, 1967; Hisel, 1974; and Yesulaitis, 1972), and various journals. Finally we gathered materials from two independent ERIC searches; these yielded 256 documents some of which led us to instruments.

As we read these sources, we were attracted by descriptions of instruments which appeared to cover one or more of the speaking and listening skills adopted by the Department of Education. Information on well over 200 instruments was examined. From this preliminary review we ordered eighty-nine instruments for more intensive study. We received and abstracted fifty-six. We also abstracted fifteen instruments that were ordered but not received. These abstracts, noted by an asterisk (*), are based on our reading of secondary sources, i.e., previously published reviews.

Our review, then, covers seventy-one instruments. We were able to obtain complete information on many instruments because we received the full instrument with all of its supporting data. However, less complete information was available for other instruments. In some cases technical data were missing. In other cases, we received only descriptions of instruments or review specimens rather than full tests or test batteries.

Method of Review

We reviewed each instrument by completing an abstract form and a rating scale.

Abstract Form

Figure 1 shows our instrument abstract form. It was designed to provide common information and to allow us to compare each instrument with our criteria.

The abstract form consists of two major parts. The first part calls for basic information (e.g., title, category, date of construction, age range, availability, and purpose) and information pertinent to our criteria. All of the items on this first part are self-explanatory except "Category." We explain how tests were designated as "speaking," "listening," or "speaking and listening" under "Method of Analysis."

The second major part consists of a checklist of the speaking and listening skills. This part called for a detailed analysis of the test items. To complete this part, we had to match test items with our skills. Where instruments did not contain separate test items, we had to infer the skills covered.

Thus, by completing an abstract form, we assembled comparable information pertinent to our criteria, including a detailed analysis of the match between our skills and particular items of a given instrument.

Figure 1

INSTRUMENT ABSTRACT FORM

(Criteria # in parentheses)

Title:

Category: _____Speaking _____Listening

Date of Construction:

Age Range:

Available from:

Purpose of Instrument:

Does the instrument clearly distinguish speaking and listening skills from other skills and/or variables (for example, reading, writing, anxiety, attitudes, knowledge, etc.)? (#'s 1 and 2)

Does the instrument focus on the presence or absence of a speaking or listening skill? (#4)

Is the instrument free from sexual, cultural, racial, or ethnic bias? (#3)

Are the situations and/or foils in the instrument suitable for the target population, clearly understandable, and allow for an acceptable range of responses? (#'s 5, 6, 7, and 15)

(Figure 1 continued)

Are procedural instructions detailed enough so that student responses are not dependent upon administrator skills in administering the instrument? (#11)

Physical description of testing procedure including time required, cost, equipment necessary, and stress factors. (#'s 12, 13, and 14)

Type of test: _____Criterion referenced ____Norm referenced

Normative data:

Validity data (#9 and 10):

Reliability data (#8):

Related documents and bibliography:

(Figure 1 continued)

LISTENING BASIC SKILLS CHECKLIST

Indicate the item numbers that correspond to the listening skills listed below.

Basic Listening Skills

1. Recognize words and phrases used by the speaker:

2. Indicate why the speaker can or cannot be understood:

Understanding What You Hear

3. Understand spoken words and ideas:

4. Identify and understand main ideas:

5. Associate important details with main ideas:

6. Understand description of events and experiences:

7. Understand speaker's purpose:

Using What You Hear

8. Understand and respond to survival words in emergency situations:

9. Summarize information and draw conclusions:

10. Recognize when words and phrases are used to convince or persuade:

11. Follow straightforward directions:

(Figure 1 continued)

SPEAKING BASIC SKILLS CHECKLIST

Indicate the item numbers that correspond to the speaking skills listed below.

Basic Oral Communication Skills

1. Use vocabulary and phrases appropriate to the situation:
2. Speak loudly enough to be heard by a listener or group of listeners:
3. Speak at rate listener can understand:
4. Say words distinctly:

Planning, Developing, and Stating Spoken Language

5. Use words in an order that clearly expresses the thought:
6. Organize main idea for presentation:
7. State main ideas clearly:
8. Support main ideas with important details:
9. Demonstrate knowledge of standard English usage:

Common Uses of Spoken Messages

10. Use survival words to cope with emergency situations:
11. Speak so listener understands purpose:
12. Ask for and give straightforward information:
13. Describe objects, events, and experiences:
14. Question other's viewpoint:

Rating Scale

Figure 2 shows our rating scale. The scale calls for dichotomous ratings (yes - no) for criteria one through four; four interval ratings (0/low to 3/high) for criteria five through nine and eleven through fifteen; and varying interval ratings for criterion ten. Ratings on criterion ten correspond with the number of different skills assessed by an instrument. For example, if an instrument assessed four speaking skills on our list, it received a rating of "4." A different instrument that assessed eight speaking skills received a rating of "8."

The purpose of the rating scale was to arrive at a recommendation to approve an instrument. This recommendation could then be used by the Department of Education.

The standards for recommending approval of an instrument are indicated by the "Summary Rating" in Figure 2. The project staff concluded that instruments should meet criteria one through four at the 100% level, criteria five through nine and eleven through fifteen at the 60% level, and criterion ten at the 70% level. These standards may appear to be somewhat arbitrary, but in determining them we were guided by three questions: (1) Do these standards reflect the relative importance of our criteria? (2) Do these standards maintain the integrity of Massachusetts assessment purposes and focus? (3) Do these standards maintain the integrity of the instrument's assessment purposes and focus?

The implication of these last two questions is important. When we indicate an instrument does not meet our standards, one should not infer the instrument is poor. First, our recommendation reflects the information we had at hand. But more important, even with complete information, we found many well conceived and developed instruments we could not recommend. These instruments were designed for different purposes from our own and measured either a broader or narrower set of skills. To recommend approval of these instruments is to abuse their purposes as well as our own. Our recommendation, then, simply says: Given the information we had at hand, given our purposes, and given our need to assess a predetermined set of skills, these instruments are/are not appropriate.

Method of Analysis

Information we gathered from our abstracts and ratings is analyzed by instrument category, level, and type. Figure 3 lists the seventy-one instruments that were reviewed. They are listed by category and level.

Category

"Category" refers to the type of skills that are assessed by a given test. The reference point for categorizing an instrument

Figure 2

INSTRUMENT RATING SCALE

Criteria 1 through 4 are evaluated on a dichotomous scale. Circle "yes" or "no" for each criterion. Criteria 5 through 9 and 11 through 15 are evaluated on a scale of 0 (-) to 3 (+). Indicate the score for each criterion in the space provided. Criterion 10 is evaluated on a variable scale that corresponds with the competencies as follows: 0 to 14 for speaking tests; 0-11 for listening tests; 0-25 for tests that assess both speaking and listening. Thus, for each of these kinds of tests determine the number of competencies that are covered and enter that number in the space provided to the left. To the right, indicate the specific test item numbers that match the competencies.

RATING			CRITERIA
Yes	No	1.	Stimulus materials require respondent to perform as a speaker/listener.
Yes	No	2.	Instrument distinguishes speaking and listening performance from reading and writing ability.
Yes	No	3.	Instrument is free of sexual, cultural, racial, ethnic content and/or stereotyping.
Yes	No	4.	Instrument assesses presence/absence of skills.

Earned Points	Possible Points		
______	(3)	5.	Instrument emphasizes application of skills in familiar situations.
______	(3)	6.	Instrument allows for optional settings.
______	(3)	7.	Instrument allows for a range of responses.
______	(3)	8.	Instrument is reliable.
______	(3)	9.	Instrument demonstrates concurrent validity.
	(3)	10.	Instrument demonstrates content validity.
______	(14)		Speaking (Note test item numbers):
______	(11)		Listening (Note test item numbers):
______	(25)		Integrative (Note test item numbers):
______	(3)	11.	Administration procedures are detailed and standardized.
______	(3)	12.	Instrument is not likely to induce undue stress.
______	(3)	13.	Instrument is cost and time effective.
______	(3)	14.	Instrument involves simple equipment.
______	(3)	15.	Instrument is suitable for our target population.

To recommend approval, instrument should receive: (a) a "yes"

(Figure 2 continued)

rating on each criterion, 1 through 4; (b) no fewer than a total of 18 points on criteria 5 through 9 and 11 through 15; (c) no fewer than 10 points for a speaking test or 8 points for a listening test or 18 points for a speaking and listening test on criterion 10.

SUMMARY RATING

a. Criteria 1 through 4 (circle one) Yes No
b. Criteria 5 through 9 and 11 through 15 (enter total points) ________
c. Criterion 10(enter total points for appropriate test)

Speaking ________
Listening ________
Both ________

Check One: Instrument approval recommended ________

Instrument approval not recommended ________ was the speaking and listening list of skills in the Basic Skills Improvement Policy. To be classified as "Speaking," an instrument had to measure one or more skills on the speaking list. Similarly, "Listening" instruments measure one or more listening skills.

Instruments which measure one or more skills from each of the speaking and listening lists were classified as "Speaking and Listening." Instruments in this category assess both speaking and listening skills in one of two ways. Some instruments measure speaking and listening separately; they contain sections and items that yield separate scores. Other measures integrate assessment by yielding a single score for both speaking and listening.

Level

"Level" refers to the age level or grade range for which the instrument was designed. Elementary school instruments are suitable for grades K - 6; secondary school instruments for grades 7 - 12. Instruments that span K - 12 are designated "Elementary and Secondary." A few tests that appeared to contain model features were designed primarily for the post secondary level and are marked with a plus (+) in Figure 3. These instruments are listed under the secondary level column because they approach that level most closely.

Of the instruments reviewed, thirty-four cover one or more speaking skills; twenty-one cover one or more listening skills; and sixteen cover both speaking and listening skills. Twenty-one of the instruments measure skills at the elementary school level; twenty-six at the secondary school level; and twenty-four at both the elementary and secondary school levels.

Instrument Type

Our instruments represent an array of measures. Some focus entirely on speaking and listening skills, while others measure a few speaking or listening skills as part of a more global behavior. For example, some standardized achievement examinations subordinate speaking skills under English expression, psycholinguistic ability, verbal knowledge, and language development.

Figure 4 depicts our instrument array another way. It shows a continuum of instrument types. Some instruments are obtrusive in that the setting in which they are utilized exists solely for the purpose of testing. A special time and place must be scheduled for these measures to be administered. Frequently, students are unfamiliar with the stimulus materials they encounter; they have never seen these particular materials before, and they must respond under carefully controlled conditions.

Other instruments are naturalistic in that they are used in settings which occur for purposes other than testing. These instruments are employed in the course of familiar, sometimes routine, activities. Any controls that exist impact on the

Figure 3

Instruments Reviewed by Category and Level

Category	Level		
	Elementary	Secondary	Elementary and Secondary
Speaking	1. Assessment of Children's Language Comprehension* 2. Classroom Communication Observation Categories 3. Observation of Socialization Behavior 4. Oral Language Evaluation 5. Roletaking Questionnaire* 6. Situational Language Tasks 7. Verbal Interaction Category System 8. Verbal Pupil-Behavior Category System* 9. ZIP Test*	10. California Achievement Tests: Forms C & D, Levels 14-19 11. Carkhuff Rating Scales+ 12. Cognitive Operations Monitored in the Classroom 13. College Outcome Measures Project+ 14. Content Evaluation Series: Language Ability Test 15. Cooperative English Test: English Expression 16. Differential Aptitude Tests: Verbal Reasoning and Language Usage 17. Essentials of English Tests 18. High School Placement Test* 19. Hill Interaction Matrix* 20. Hoyum-Sanders Junior High School English Test	27. California Achievement Tests: Form A, Levels 2-5 28. Cognitive Abilities Test: Verbal Battery 29. Dramatic Acting Test* 30. Family Interaction Scales 31. Florida Taxonomy of Cognitive Behavior 32. Science Research Associates Assessment Survey Achievement Series* 33. Test of Ability to Explain 34. Tests of Adult Basic Education: Language Expression

Figure 3 (continued)

Category	Level		
	Elementary	Secondary	Elementary and Secondary
Speaking (continued)		21. Iowa Tests of Educational Development* 22. McRel Interaction Analysis System 23. Minnesota High School Achievement Examinations: Language Arts 24. Missouri College English Test* 25. Resource Process Coding System*+ 26. Smith System-Strategies of Teaching+	
Listening	35. Cooperative Primary Tests: Listening 36. Durrell-Sullivan Reading Capacity and Achievement Tests 37. Flowers Auditory Test of Selective Attention* 38. Murphy-Durrell Reading Readiness Analysis 39. Test for Auditory Comprehension of Language 40. Test of Listening Comprehension*	42. Brown-Carlsen Listening Test 43. Cloze Listening Test 44. Effective-Listening--Pretest and Posttest 45. Giving and Taking Directions 46. Jones-Mohr Listening Test*+	47. Auditory Skills Test Battery 48. Durrell Listening--Reading Series 49. Fundamental Achievement Series: Verbal 50. Learning Through Listening 51. Michigan Educational Assessment Program Listening Test 52. Peabody Picture Vocabulary Test 53. STEP Listening Test 54. Test of Auditory Discrimination 55. Utah Test of Language Development

Figure 3 (continued)

Category	Level		
	Elementary	Secondary	Elementary and Secondary
Listening (continued)	41. Tests of Basic Experiences, Second Edition, Language		
Speaking and Listening	56. Bowen Language Behavior Inventory* 57. California Achievement Tests, Form A, Level 1 58. Illinois Test of Psycholinguistic Abilities 59. Interactional Competency Checklist 60. Language Communication Skills Task	61. Functional Communication Profile+ 62. Ilyn Oral Interview 63. NAEP Speaking and Listening Field Tests 64. Palo Alto Group Therapy Scales+	65. Comprehensive Test of Basic Skills, Form S, Levels A,C,2, & 4 66. English Language Testing Project for the Bureau of Indian Affairs 67. Flanders System of Interaction Analysis 68. Introduction to Speech Communication 69. Metropolitan Achievement Tests 70. School and College Ability Tests: SCAT Series II 71. Short Form Test of Academic Aptitude: Vocabulary and Memory Tests

* Review based on secondary source
+ Post secondary level measure

observer using the instrument, not on the student(s) being observed. For example, the observer may need to be trained in coding various types of behavior, but the student being observed is not constrained to behave one way or another.

Some instruments are evaluative in that the information they produce is used to identify individual strengths and weaknesses, to compare students, to place persons in courses, curriculums and jobs, to make judgments about the achievement of institutions that teach students, and to determine proficiency levels.

Other instruments are descriptive. They tell us the kind and amount of behavior that is observed in a given setting, but they do not determine levels of proficiency or make other evaluations.

The instruments we reviewed can be classified generally as formal tests, interaction observation systems, and functional performance ratings. Formal tests are characterized as obtrusive and evaluative; interaction observation systems as naturalistic and descriptive.

Between these ends are functional performance ratings. These instruments are naturalistic and evaluative. They employ settings that are familiar to students -- telling stories, asking and answering questions, interviewing, conversing with others -- but they produce information that identifies a student's level of proficiency.

Figure 4 shows where each instruments fit on our continuum, i.e., which are formal tests, which are functional performance ratings and which are interaction observation systems. These classifications pertain to the general characteristics of the particular instruments we reviewed; they are not meant to be descriptive of all assessment instruments.

We emphasize that the typology described in Figure 4 was developed during our analysis of instruments. This typology was not predetermined; it did not guide us in selecting instruments for review.

We use the term "formal test" rather than "standardized test" because at least five measures in this group are not yet standardized. However, these instruments do have the other characteristics of standardized tests.

Findings: Instrument Adoption

Abstracts and ratings for each of the instruments reviewed are included in Appendix A. In this chapter we summarize our findings through a series of tables that were developed from the instrument rating scales and the skills checklists. First we examine how well our instruments meet criteria one through nine and eleven through fifteen. Next, we will see how well they meet

FIGURE 4

INSTRUMENT TYPES

Obtrusive Evaluative	Naturalistic Evaluative	Naturalistic Descriptive
Formal Tests	Functional Performance Ratings	Interaction Observation Systems
1, 10, 14-18, 20, 21 , 23, 24, 27, 28, 32, 34-44, 46-54, 56-58, 63, 65, 68-71*	4-6, 9, 11, 13, 29, 33, 45, 55, 59-62, 66	2, 3, 7, 8, 12, 19, 22, 25, 26, 30, 31, 64, 67

*See Figure 3 for names of instruments.

criterion ten. Then, we will draw conclusions regarding instrument adoption.

Criteria 1 - 9 and 11 - 15

Tables 1, 2, and 3 summarize criteria one through nine and eleven through fifteen ratings for speaking, listening, and speaking and listening instruments respectively.

An examination of the ratings for instrument number one in Table 1 will help in reading the tables. Instrument number one received a "No" rating on criteria 1, 2, and 4, and a "Yes" rating on criterion 3. Since it did not receive a "Yes" rating on all these criteria, the summary rating for 1 through 4 is "No."

The ratings for criteria 5 - 9 and 11 - 15 range from 0 for criteria 7 and 15 to 3 for criterion 8. The total of these ratings is 14.

The standards for recommending approval of an instrument -- exclusive of criterion 10 -- call for a summary rating of "Yes" on criteria 1 - 4 and a total of at least 18 points for criteria 5 - 9 and 11 - 15. We can see that instrument number one does not meet these standards. From this kind of reading, we can see how each instrument compares with two of our standards for instrument approval. One additional note is in order. Blank spaces for some instrument ratings signify insufficient information was available to make a judgment.

Table 1 --speaking instruments-- indicates that two elementary school instruments, #2 and #6, meet two of our three standards. On the secondary school level, two of our standards are met by three instruments: #11, #13, and #22. None of the instruments for both elementary and secondary schools meet the two standards.

Table 2 --listening instruments-- shows that four instruments meet the two standards; #43 and #45 are for secondary school students; and #51 and #54 are for both elementary and secondary school students.

Table 3 --speaking and listening instruments-- indicates that three instruments meet the two standards: #61 and #64 are for the secondary school level; and #67 is for both the elementary and secondary school levels.

So far, then, we have the following instruments that partially meet our standards:

Table 1. Speaking Instrument Ratings for Criteria 1 through 9 and 11 through 15

Level and Instrument Number	CRITERIA*														Summary 1-4	Rating 5-9+ 11-15
	CONTENT							TECHNICAL								
	1	2	3	4	5	6	7	8	9	11	12	13	14	15		
Elementary																
1	N	N	Y	N	2	1	0	3	1	2	2	1	2	0	N	14
2	Y	Y	Y	Y	3	3	2	0	0	2	3	3	3	2	Y	21
3	Y	Y	Y	N	2	1	1	2	0	1	1	1	0	1	N	10
4	Y	Y	Y	N	3	1	2			2	3	1	3	1	N	16
5	N	N		N	2	1	1	2	3	2	2	3	3	2	N	21
6	Y	Y	Y	Y	3	3	2			2	3	2	3	1	Y	19
7	Y	Y	Y	N	2	2	3	1	1	2	3	3	3	1	N	21
8	Y	Y	Y	Y	2	1	1	0	0	2	3	3	3	2	Y	17
9	Y	Y		N	2	1	3	3	3	2	2	1	3	1	N	21
Secondary																
10	N	N	Y	N	1	0	1			3	2	3	2	3	N	15
11	Y	Y	Y	Y	1	1	3	3	3	2	1	1	2	1	Y	18
12	Y	Y	Y	N	2	2	1	0	0	2	3	3	1	1	N	15
13	Y	Y	Y	Y	3	3	3	3	3	2	2	2	2	0	Y	23
14	N	N	Y	N	2	1	1			2	2	3	3	3	N	17
15	N	N	Y	N	2	1	1	3	3	3	2	3	3	2	N	23
16	N	N	Y	N	1	1	1	3	3	3	2	2	3	3	N	22
17	N	N	Y	N	2	1	1	3	2	3	2	3	3	3	N	23
18	N	N		N	1	1	1			3	2	2	3	3	N	16
19	Y	Y	Y	Y	3	2	2	0	0	2	3	3	2	0	Y	17
20	N	N		N	0	0	1	3	2	2	1	3	3	3	N	18
21	N	N		N	2	1	1			3	2	2	2	2	N	15
22	Y	Y	Y	Y	3	2	2	0	0	2	2	3	3	2	Y	19
23	N	N	Y	N	1	1	1		2	3	2	3	3	3	N	19
24	N	N	Y	N	3	1	1	3	3	3	2	3	3	1	N	23
25	Y	Y	Y	N	3	2	2	0	0	2	2	2	2	1	N	16
26	Y	Y	Y	Y	2	2	2	1	1	1	2	1	0	1	Y	13

Table 1 (continued)

Speaking Instrument Ratings for Criteria 1 through 9 and 11 through 15

Level and Instrument Number	CRITERIA* CONTENT 1	2	3	4	5	6	7	TECHNICAL 8	9	11	12	13	14	15	Summary 1-4	Rating 5-9+ 11-15
Elementary and Secondary																
27	N	N	Y	N	1	0	1	2	3	3	2	3	2	3	N	20
28	N	N	Y	N	1	1	1	3	3	3	2	3	3	3	N	23
29	Y	Y	N	Y	3	3	3	3		2	2	2	3	3	N	24
30	Y	Y	Y	Y	2	2	2	0	0	2	1	1	0	1	Y	11
31	Y	Y	Y	N	3	3	2	0	0	1	2	3	3	1	N	18
32	N	N	?	N	1	1	1			3	2	3	3	2	N	16
33	N	N	Y	N	1	1	1	2	3	3	2	3	3	2	N	21
34	N	N	Y	N	1	1	1			3	2	3	3	3	N	17
Average Ratings (N=34)	17N;17Y	17N;17Y	1N;6?;27Y	24N;10Y	1.94	1.41	1.53	1.26	1.55	2.29	2.09	2.38	2.44	1.79	25N; 9Y	18.29

*Criteria 1-4 are rated on a dichotomous scale. Y = yes; N = no.
Criteria 5-9 and 11-15 are rated on a scale ranging from 0 (low) to 3 (high).

Table 2. Listening Instrument Ratings for Criteria 1 through 9 and 11 through 15

Level and Instrument Number	CRITERIA* CONTENT							TECHNICAL							Summary 1-4	Rating 5-9+ 11-15
	1	2	3	4	5	6	7	8	9	11	12	13	14	15		
Elementary																
35	Y	Y	Y	N	2	1	1	3	3	3	2	3	3	2	N	23
36	Y	N	Y	N	1	1	1	3	3	2	2	3	3	2	N	21
37	Y	Y		N	2	1	1	1	1	1	2	2	2	2	N	15
38	Y	N	Y	N	1	1	1	3	3	2	2	3	3	1	N	20
39	Y	Y	Y	Y	2	2	0	3	3	2	1	1	2	0	Y	16
40	Y	Y	Y	Y	2	1	1	3	3		2	2	1	2	Y	17
41	N	N	Y	N	2	1	1			2	2	3	3	2	N	16
Secondary																
42	Y	N	Y	Y	2	1	1	2	2	3	1	2	2	1	N	17
43	Y	Y	Y	Y	1	1	1	3	2	3	1	2	3	1	Y	18
44	Y	Y	Y	Y	2	2	2	0	0	3	2	1	2	1	Y	15
45	Y	Y	Y	Y	3	1	2	1	1	2	2	1	2	3	Y	18
46	Y	N	Y	Y	1	2	1	1	1	2	1	1	1	1	N	12
Elementary and Secondary																
47	Y	Y	Y	N	1	1	1	3	3	3	2	1	2	2	N	19
48	Y	N	N	Y	1	1	1	3	3	1	2	2	3	3	N	20
49	Y	N	N	N	3	2	1	3	3	3	2	3	3	3	N	26
50	Y	Y	N	Y	3	2	1	3	3	3	2	3	2	1	N	23
51	Y	Y	Y	Y	3	2	1			3	3	3	3	3	Y	21
52	Y	Y	Y	N	1	1	0	3	3	2	2	1	3	2	N	18
53	Y	N	Y	Y	1	1	1	2	3	3	2	2	2	2	N	19
54	Y	Y	Y	Y	2	1	1	2	3	3	3	2	2	3	Y	22

Table 2 (continued)

Listening Instrument Ratings for Criteria 1 through 9 and 11 through 15

Level and Instrument Number	CRITERIA* CONTENT 1	2	3	4	5	6	7	TECHNICAL 8	9	11	12	13	14	15	Summary 1-4	Rating 5-9+
Elementary and Secondary (cont)																
55	Y	Y	Y	Y	2	2	2	3	3	3	1	0	1	2	Y	17
Average Ratings (N=21)	20Y;1N	13Y;8N	17Y;1?;3N	13Y;8N	1.81	1.33	1.05	2.19	2.19	2.33	1.86	1.95	2.29	1.86	8Y;13N	18.71

*Criteria 1-4 are rated on a dichotomous scale; Y = yes; N = no.
Criteria 5-9 and 11-15 are rated on a scale ranging from 0 (low) to 3 (high).

Table 3. Speaking and Listening Instrument Ratings for Criteria 1 through 9 and 11 through 15

Level and Instrument Number	CRITERIA*														Summary 1-4	Rating 5-9+ 11-15
	CONTENT									TECHNICAL						
	1	2	3	4	5	6	7	8	9	11	12	13	14	15		
Elementary																
56																
57	N	N	Y	N	1	0	1	3	3	2	1	3	2	2	N	18
58	Y	Y	Y	N	2	0	1	3	3	3	2	1	1	1	Y	17
59	Y	Y	Y	Y	3	3	3			1	1	1	1	1	Y	14
60	Y	Y	Y	Y	3	0	3	2	2	2	2	1	2	0	Y	17
Secondary																
61	Y	Y	Y	Y	3	2	3	3	3	2	3	1	3	1	Y	24
62	Y	Y	Y	N	3	1	2	3	3	3	3	1	2	1	N	22
63	N	N	Y	N	2	0	1	2	3	2	1	2	3	2	N	18
64	Y	Y	Y	Y	2	2	3	2	2	3	3	3	3	1	Y	24
Elementary and Secondary																
65	Y/N	N	Y	N	2	1	1			3	2	3	3	3	N	18
66	Y	Y	Y	Y	1	1	3	0	0	0	2	1	1	0	Y	9
67	Y	Y	Y	Y	2	2	2	2	2	3	3	3	3	1	Y	23
68	N	N	Y	N											N	
69	Y/N	N	Y	N	2	1	1	3	3	2	2	3	3	3	N	23
70	N	N	Y	N	1	1	1	3	3	3	2	3	3	2	N	22
71	N	N	Y	N	2	1	1	3	3	2	2	3	3	3	N	23
Average Ratings (N=15)**	8Y;5N;2YN	8Y;7N	15Y	6Y;9N	1.93	1.0	1.73	1.93	2.0	2.07	1.93	1.93	2.20	1.40	7Y;8N	18.13

*Criteria 1-4 are rated on a dichotomous scale; Y = yes, N = no.
Criteria 5-9 and 11-15 are rated on a scale ranging from 0 (low) to 3 (high).
**This N excluded #56.

Speaking: Elementary
#2 - Classroom Communication Observation Categories
#6 - Situational Language Tasks

Secondary
#11- Carkhuff Rating Scales
#13- College Outcome Measures Project
#22- McRel Interaction Analysis System

Listening: Secondary
#43- Cloze Listening Test
#45- Giving and Taking Directions

Elementary and Secondary
#51- Michigan Educational Assessment Program
#54- Test of Auditory Discrimination

Speaking and Listening: Secondary
#61- Functional Communication Profile
#64- Palo Alto Group Therapy Scales

Elementary and Secondary
#67- Flanders System of Interaction Analysis

Criterion 10

Tables 4, 5, and 6 summarize the skills that are stressed in speaking, listening, and speaking and listening instruments respectively. These tables show the particular skills and the total number of skills that are assessed by each instrument. For example, in Table 4, an "x" appears in columns 12 and 14 for instrument #2. This means only two skills -- #12 - asking for and giving straightforward information and #14 - questioning other's viewpoint -- are assessed by instrument #2.

Our standards for recommending approval of an instrument for criterion ten require ten skills to be assessed by speaking instruments, eight skills by listening instruments, and eighteen skills by both speaking and listening instruments. In Table 4 we can see that only instrument #13 qualifies as a measure of speaking. Table 5 shows only instrument #51 qualifies as a measure of listening. Table 6 indicates none of the instruments that assess both speaking and listening qualifies as a measure.

Conclusions

When we combine our findings for criteria one through nine and eleven through fifteen with our findings for criterion ten, we find two instruments meeting all of the standards. One is a speaking instrument -- #13 College Outcome Measures Project -- and one is a listening instrument -- #51 Michigan Educational Assessment Program Listening Test.

After reviewing these instruments again, we concluded we

Table 4. Skills Stressed in Speaking Instruments (Criterion 10)

Level and Instrument Number	Speaking Skills														Total Skills
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Elementary															
1															
2												x		x	2
3															
4	x				x		x	x	x				x		6
5															
6															
7												x			1
8									x		x	x	x	x	5
9	x				x				x						3
Secondary															
10					x				x						2
11															
12															
13	x	x	x	x	x	x	x	x	x		x	x	x	x	13
14	x								x						2
15	x								x						2
16	x				x				x						3
17	x				x				x						3
18	x				x				x						3
19															
20									x						1
21	x				x	x			x						4
22											x	x	x	x	4
23	x				x				x						3
24									x						1
25												x			1
26												x	x		2
Elementary and Secondary															
27									x						1
28	x				x										2

Table 4. (continued) Skills Stressed in Speaking Instruments (Criterion 10)

Level and Instrument Number	Speaking Skills 1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total Skills
Elementary and Secondary (continued)															
29	x	x	x		x										4
30	x	x	x												3
31						x	x	x	x			x	x		6
32	x				x	x			x						4
33							x	x				x		x	4
34					x				x						2
Total Instruments Covering Each Skill	14	3	3	1	13	4	4	4	18	0	3	9	6	5	

Table 5. Skills Stressed in Listening Instruments (Criterion 10)

Level and Instrument Number	Listening Skills 1	2	3	4	5	6	7	8	9	10	11	Total Skills
Elementary												
35	x		x			x			x	x	x	6
36	x		x	x	x	x						5
37	x											1
38	x											1
39	x		x									2
40												
41	x		x		x							3
Secondary												
42	x		x	x	x		x		x		x	7
43	x		x	x	x		x					5
44	x		x	x	x	x			x			6
45										x		1
46	x				x							2
Elementary and Secondary												
47	x											1
48	x	x										2
49	x		x								x	3
50	x		x	x	x		x		x			6
51	x		x	x	x	x	x		x	x	x	9
52	x		x									2
53					x	x	x		x	x		5
54	x											1
55	x		x						x			3
Total Instruments Covering Each Skill	18	0	13	6	9	5	5	0	7	4	4	

Table 6. Skills Stressed in Speaking and Listening Instruments (Criterion 10)

Level and Instrument Number	Speaking Skills														Listening Skills											Total Skills
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	1	2	3	4	5	6	7	8	9	10	11	
Elementary																										
56																										
57									x						x		x		x						x	5
58				x	x				x				x		x		x									6
59	x	x	x	x	x	x	x	x				x	x	x	x		x		x	x					x	16
60																										
Secondary																										
61	x				x					x		x			x		x	x	x	x	x	x			x	12
62	x	x	x	x	x				x			x	x		x		x						x		x	12
63						x	x	x			x	x	x					x	x	x	x		x	x	x	13
64	x													x			x			x	x					5
Elementary and Secondary																										
65	x			x	x	x	x	x							x		x		x	x			x			11
66					x								x		x		x									4
67																										
68																										
69									x						x		x		x	x					x	6
70	x														x		x									3
71	x																x		x	x						4
Total Instruments Covering Each Skill	7	2	2	3	6	3	3	3	5	1	1	4	5	2	9	0	11	2	7	7	3	1	3	1	6	

should not recommend them for state adoption at this time. The College Outcome Measures Project is a postsecondary level instrument. The Michigan listening test is still being developed. Hence, we cannot recommend one instrument because it clearly is not suitable for our target population, and it would simply be premature to recommend the other instrument.

Do these conclusions invalidate our criteria? No. Our criteria and standards did clearly indicate that two instruments come close to meeting our needs. Therefore, our criteria do discriminate among different tests and do indicate that an appropriate test can be constructed. The criteria represent a device for screening an array of instruments. They help to focus on instruments that meet our needs best. But in the final analysis, a decision to recommend adoption must rest with examination of an instrument's unique characteristics in light of each of the criteria. The criteria constitute the basis for our conclusion.

Findings: Feature Analysis

Unable to recommend any single instrument for use in its present form, we turned our attention to locating features that might be incorporated in one or more newly created instruments. We did this first by examining the overall "fit" of the instruments reviewed with each of the criteria developed. This fit is reflected by the average ratings in Tables 1, 2, and 3 and by the total number of instruments covering each skill in Tables 4, 5, and 6.

We will explore this fit, but first we emphasize that there is a great deal of variation among our instruments. In the following analysis we stress trends. Some of these trends -- especially those that interrelate instrument category and instrument type with each criterion -- are based on detailed tabulations. Rather than include these tabulations, we elected to report the highlights they produced. Our findings are organized around the criteria.

Criteria 1 - 4

1. Stimulus materials should require students to demonstrate skill as a speaker or listener.
2. Assessment should clearly distinguish speaking and listening performance from reading and writing ability.
3. Assessment should be free of sexual, cultural, racial, ethnic content and/or stereotyping.
4. Assessment should confirm the presence or absence of skills.

Criterion one was met by more of the listening instruments than speaking instruments. Criterion two was also met by more of the listening instruments, but to a lesser degree than criterion one. Criterion three was met relatively well by both types of instruments. Criterion four was met far better by listening instruments than by speaking instruments.

It appears that our listening instruments require actual performance and assess the presence or absence of skills more than our speaking instruments. Our listening instruments also tend to be less mediated by ability to read and write than speaking instruments, but this difference is less pronounced. In fact, eight listening tests that cover elementary and secondary school levels do tend to be mediated by reading and writing.

Why these differences between speaking and listening instruments? Part of the answer lies in the type of instrument. We have a greater array of speaking instruments (Figure 4). Eleven speaking measures are interaction observational systems, eight are functional performance ratings, and twelve are formal tests. The spread of instruments that assess both speaking and listening range from two interaction observation systems to five functional performance ratings to nine formal tests.

The speaking instruments that are interaction observational systems and functional performance ratings do in fact meet criteria one and two very well, and they also tend to meet criterion four. But the speaking instruments that are formal tests do not meet criteria one, two, or four. On the other hand, the listening instruments, predominately formal tests, tend to meet these criteria.

We can conclude that different types of instruments can be employed to assess speaking and listening directly. Interaction observation systems and functional performance ratings seem to be more suitable for assessing speaking skills, and formal tests can work well for assessing listening.

Before we go on, we need to say more about criterion three. Since most of the instruments reviewed were developed in the past decade, their content tends to be sensitive to bias and stereotyping. However, many instruments may have an evaluative point of view. Responses may be judged to be correct or appropriate by a given value system. Indeed, the concept of individual achievement testing itself may reflect a cultural value. Instruments may reflect this value system even though they avoid content that is clearly stereotypic.

Criteria 5-9 and 11 - 14

The average ratings in Tables 1, 2, and 3 allow us to compare the remaining "content" criteria -- five, six, and seven -- with the "technical" criteria -- eight, nine, and eleven through fifteen. All three categories of instruments -- speaking, listening, speaking and listening -- generally receive higher ratings for the technical criteria than for the content criteria. Again, this varies with whether the instrument is a formal test, functional performance rating, or interaction observation instrument. We will examine these criteria more.

5. Assessment should emphasize application of speaking and listening skills in familiar situations.
6. Assessment should allow for optional settings.
7. Assessment should involve situations that allow for a range of responses.

Criteria five, six, and seven are met better by functional performance ratings and interaction observation systems. The latter call for observing behavior in familiar classroom settings, and they allow for a greater range of responses than formal tests which require single correct answers under controlled conditions.

8. Assessment should demonstrate that outcomes are more than just chance evidence.
9. Assessment should provide results that are consistent with other available evidence.

Criteria eight and nine generally are met best by our formal tests. The functional performance ratings also approach these criteria better than the interaction observation instruments. Technical data were sketchy or nonexistent for interaction observation instruments. This is due primarily to the purpose of such instruments; they are designed for training and research rather than evaluation. Consequently they are not checked with other instruments. The formal tests, most of which are standardized, have a great deal of supporting data.

Embedded within our analysis of criteria eight and nine, we found one exception to the superiority of formal tests. With speaking instruments, functional performance ratings meet both criteria eight and nine slightly better than formal tests. Hence, our instruments indicate that when carefully constructed, functional performance rating scales can be reliable and valid for evaluating speaking skills.

11. Assessment procedures should be standardized and detailed enough so that individual responses will not be dependent on administrators' skills in administering the procedures.

All of the instruments require some skill on the part of the administrator, but this skill could be learned with a small amount of training. The formal tests rated better than the other types of instruments on this criterion because they are accompanied by detailed instructions for administration.

We observed, however, that formal tests of listening do not rate as well as formal tests of speaking. Many of the listening tests rely on oral reading skills of the teacher or administrator. This feature can affect student responses. To avoid this pitfall, it might be more desirable to present oral materials on audio or video tape.

12. Assessment procedures should be as free as possible of such negative outcomes as undue stress or anxiety.

With criterion twelve, the type of instrument is the critical factor in reducing anxiety. Interaction observation systems have the advantage. Since they are generally unobtrusive, they present little stress for the student. Functional performance ratings also are advantageous over formal tests.

13. Assessment should be practical in terms of cost and time.
14. Assessment should involve simple equipment such as most school districts have on hand.

Speaking instruments tend to meet these criteria more than listening instruments, but the type of instrument is also a factor. Interaction observation systems and formal tests are more efficient and economical; they also demand less in special equipment. The functional performance ratings reviewed were designed to be administered to individuals or pairs of students. Sometimes they require unique stimulus materials or special recording devices. To better meet these criteria, functional performance ratings need to be designed for use with larger groups of students in regular classroom activities.

15. Assessment should be suitable for the target population of the basic skills improvement program.

Ratings on criteria fifteen are related to the type and category of instrument. Formal tests received substantially higher ratings than functional performance ratings and interaction observation systems. This means the population for which our formal tests were designed coincides best with our own target population. Functional performance ratings we reviewed frequently were designed for special populations such as bilingual students or students with language or speech disabilities. Our interaction observation instruments, on the other hand, frequently were designed for a different age population than our own.

We also found among our formal tests that speaking instruments received higher ratings than listening instruments. Why? Because criterion fifteen has embedded within it three factors that define our target population: age, "mainstream," and recency. Recency accounts for the overall lower ratings of listening instruments. While our formal tests of listening are more suitable for the age range and "mainstream" student, more of these same tests are older measures and may contain items that were more appropriate when the test was constructed than today.

Criterion 10

10. Assessment should correspond as closely as possible with the given speaking and listening skills.

We analyzed criterion ten data first by rank ordering the emphasis on skills. This identifies the skills that are emphasized most and least in the instruments reviewed. Second, we rank ordered the instruments that most closely match our skills.

Emphasis on Skills

Speaking. Our rank ordering of skills was based on the total number of instruments that cover each skill. (See Tables 4, 5, 6.) To rank order the speaking skills, we combined the data from Table 4 with the data on speaking from Table 6.

The skills, in order of rank, are shown in Table 7. Skill #9 -- demonstrating knowledge of standard English usage -- is emphasized most. Twenty-three instruments assess this skill. Close behind are skills #1 and #5 -- using vocabulary and phrases appropriate to the situation and using words in an order that clearly expresses the thought. These language skills are related to vocabulary and general verbal knowledge.

At the other extreme, skill #10 -- using survival words to cope with emergency situations -- is emphasized least; only one instrument approaches this skill. Between these extremes are the skills related to organization and purposeful presentation of ideas.

Listening. Table 8 shows the rank order of listening skills. Information in this table was derived by combining data from Table 5 with the data on listening in Table 6.

Skill #1 -- recognizing words and phrases used by the speaker -- is emphasized most. A substantial majority of the instruments reviewed assess this skill. Close behind it is skill #3 -- understanding spoken words and ideas. Listening instruments, like speaking instruments, assess most those skills that are related to vocabulary and general verbal knowledge.

Skill #2 -- indicating why a speaker can or cannot be understood -- is not assessed by any instrument. This skill calls for listeners to detect problems associated with attending to oral messages -- being able to hear the speaker, following the speaker's rate, recognizing distinctness. We do not know why the listening instruments omit this skill. Three possible reasons are: (a) The skill is ill conceived. (b) Test publishers do not consider the skill a significant variable for assessing listening; that is, this skill does not fit conceptions of listening as measured by tests. (c) Test publishers have been unable to devise means for assessing the skill.

We believe the second explanation is most plausible. We found that many listening instruments go beyond our skills to include drawing inferences and making judgments based upon spoken messages. Moreover, some items which test these skills could be answered without listening to the stimulus materials. Listening instruments, then, assess critical thinking more than

Table 7

RANK ORDER OF SPEAKING SKILLS

Rank	Skill Number	Number of Instruments (N=50)
1	9	23
2	1	21
3	5	19
4	12	13
5	13	11
6	6, 7, 8, 14	7
7	2, 3	5
8	4, 11	4
9	10	1

Table 8

RANK ORDER OF LISTENING SKILLS

Rank	Skill Number	Number of Instruments (N=37)
1	1	27
2	3	24
3	5	16
4	6	16
5	9, 11	10
6	4, 7	8
7	10	5
8	8	1
9	2	0

detecting attributes of speaking that impede or enhance attending to, understanding, and responding to the message.

A second skill that is emphasized least is skill #8 -- understanding and responding to survival words in emergency situations. This is consistent with what we found in speaking instruments. Apparently using and responding to survival messages are unique skills in our program.

Between these extremes of most and least emphasized skills are those that include comprehending the organization and purposes of spoken messages.

Instrument - Skills Correspondence

We based our rank ordering of instruments on the total number of different skills covered by each instrument. (See Tables 4, 5, and 6). Again we combined information from Tables 4 and 5 with information from Table 6.

Three speaking instruments covered fifty percent -- seven out of fourteen -- or more of our skills. In order of rank, they are:

1. #13 - College Outcome Measures Project (covers 13 skills)
2. #59 - Interactional Competency Checklist (covers 11 skills)
3. #62 - Ilyn Oral Interview (covers 8 skills)

Seven listening instruments covered fifty percent -- six out of eleven -- or more of our skills. In order of rank, they are:

1. #51 - Michigan Educational Assessment Program Listening Test (covers 9 skills)
2. #61 - Functional Communication Profile (covers 8 skills)
3. #42 - Brown-Carlsen Listening Test (covers 7 skills)
   #63 - NAEP Speaking and Listening Field Tests (covers 7 skills)
4. #35 - Cooperative Primary Tests: Listening (covers 6 skills)
   #44 - Effective Listening: Pretest and Posttest (covers 6 skills)
   #50 - Learning Through Listening (covers 6 skills)

Table 9 charts these measures by category and level. We regarded these instruments as potential models for developing new instruments.

Conclusions

Our last task was to compare each of these instruments again with the other criteria, one through nine and eleven through

Table 9

POTENTIAL MODEL INSTRUMENTS THAT ASSESS FIFTY PER CENT OR MORE OF THE SPEAKING AND LISTENING SKILLS BY CATEGORY AND LEVEL

CATEGORY	LEVEL	
	Elementary	Secondary
Speaking	#59 - Interactional Competency Checklist	#13 - College Outcome Measures Project+ #62 - Ilyn Oral Interview
Listening	#51 - Michigan Educational Assessment Program: Listening Test #35 - Cooperative Primary Test #50 - Learning Through Listening	#51 - Michigan Educational Assessment Program: Listening Test #61 - Functional Communication Profile+ #42 - Brown-Carlsen Listening Test #63 - NAEP Field Test (Listening) #44 - Effective Listening: Pretest and Posttest #50 - Learning Through Listening

fifteen. We eliminated instruments #59, #62, #42, and #35 for the following reasons: (1) Instrument #59 failed to meet our standards for criteria 5-9 and 11-15. This instrument was developed primarily for research and is just now being adapted to classroom use. Since it requires videotaping students in sociodramatic play and analyzing the tapes in detail, it is too expensive in terms of cost and time. (2) Instruments #62, #42, and #35 fail to meet two of the three standards.

Instruments #63, #44, and #50 also failed to meet two of the three standards, but we decided to include these in our final list of instruments with model features because of the unusual conditions that yield their ratings.

Instrument #63 - NAEP Field Tests in Speaking and Listening -- contains two separate measures, one for speaking and one for listening. We rated this test as one instrument. When we separated the two measures we found our ratings of the speaking portion contaminated our ratings of the listening portion. The speaking test does not assess skill directly. The listening test does. Moreover, the listening test contains some well developed stimulus materials. We concluded this test should stay on our list of model instruments as a listening measure.

Instrument #44 - Effective Listening: Pretest and Posttest - is a nonstandardized pretest and posttest that accompanies a listening training program. The instrument meets our first four criteria, the most crucial ones. Like the NAEP listening test, it contains good stimulus materials that might be models for future instruments.

Instrument #50 - Learning through Listening - was developed in Great Britain. The accents and vocabulary students will hear are distracting (e.g., "Can you get this bloke's clobber up to Room 209 please?"). But the kinds of stimulus materials used are excellent; they should be studied by future developers of listening instruments.

We therefore include six instruments in our final list:

#13 - College Outcome Measures Project
#51 - Michigan Educational Assessment Program Listening Test
#61 - Functional Communication Profile
#63 - NAEP Field Test (Listening)
#44 - Effective Listening: Pretest and Posttest
#50 - Learning Through Listening

None of these instruments is an interaction observation system. Two are functional performance ratings: 13 assesses speaking; 61 assesses speaking and listening (but the listening component meets our criteria better than the speaking component). The remaining instruments are formal tests that assess listening.

None of these instruments is suitable for adoption by the state at this time. But they do contain features we incorporate in our recommendations (see Chapter IV). We suggest that the Department of Education, local school districts and individuals or firms that contemplate developing new instruments consult these existing instruments.

SUMMARY

The state of the art of speaking and listening assessment derives from investigations of speech rating scales, functional communication measures, listening tests, and oral language measures for bilingual students. This research has yielded a variety of measures. Among the instruments we reviewed are formal tests, functional performance ratings and interaction observational systems.

When assessing speaking and listening, one needs to consider an array of measuring devices. Too often assessment is interpreted to mean "testing" in a narrow sense, i.e., using only published standardized tests. Such tests have value but represent only one kind of instrument that might be used.

The instruments we reviewed indicate listening is assessed primarily through formal tests. Speaking, however, is assessed through other measures. When developing new instruments, there appears to be no compelling reason for assessing speaking and listening in the same way. Each complex of skills can be assessed with different types of instruments.

Of the seventy-one instruments reviewed, none meet our predetermined criteria for adoption. New instruments tailored to the purposes of the Basic Skills Improvement Policy and to distinct characteristics of speaking and listening behavior must be developed.

Of the seventy-one instruments reviewed, six contain features that ought to be incorporated in new instruments. These features are described in Chapter 4.

CHAPTER III

STATE OF THE ART OF ASSESSMENT OF SPEAKING AND LISTENING SKILLS IN OTHER STATES AND THEIR JURISDICTIONS

Scope

This chapter reports speaking and listening assessment practices in other states and local jurisdictions. Data were obtained from a number of sources: selected manuscripts found among the 256 documents yielded by two independent ERIC searches; national reports of the Educational Testing Service, the National Association of Secondary School Principals, the Education Commission of the States, and the National Consortium on Testing; personal communication with state and local officials connected with minimal competency and basic skills programs in speaking and listening; and state and local school district curriculum publications.

State and local activities in basic skills and minimal competency programs change so rapidly it is difficult to be certain that we have reported all practices related to the assessment of speaking and listening. However, we did attempt to verify the existence of speaking/listening programs with Dr. Chris Pipho, Associate Director of the Department of Research and Information of the Education Commission of the States, and with Mary Tobin, Research Associate with National Evaluation Systems. Dr. Pipho regularly reports on state activity in minimal competency testing, and Mary Tobin is currently engaged in a four year study of minimal competency testing programs. Both individuals indicated that programs in speaking and listening are not as prevalent as programs in reading, writing and mathematics, and that the programs reported here constitute the major activity in speaking/listening.

This section of the report accomplishes three purposes: (1) It reviews current practices, first in a general way, then state-by-state and jurisdiction-by-jurisdiction. (2) It summarizes practices found throughout the country. (3) It compares prevalent practices with the criteria reported in Chapter I.

Review of Practices

General Trends

In the past decade, states and local school districts have been responding to the demand for educational accountability in at least three general ways: by implementing statewide norm-referenced testing programs; by implementing statewide assessment programs that include a variety of measurement; and by implementing statewide and local minimum competency programs.

With statewide testing programs, norm referenced tests report student scores by comparing a student with other students, i.e., by determining an individual's relative standing in some norm group. Standardized tests are basic to statewide and local norm-referenced testing programs.

In statewide assessment programs, standardized norm-referenced tests may be a part of some assessment. Criterion referenced tests are also frequently a part of this assessment; these tests tell what an individual can or cannot do with regard to predetermined objectives or requirements.

Minimum competency programs are designed to assess whether students have acquired some specific set of skills at a predetermined level. Such programs frequently have a cut-off point that differentiates between students who have achieved minimum competencies and those who have not. Such programs frequently, but not always, use minimum competency achievement to identify students who need remedial instruction, to make grade promotion decisions, and to graduate students.

Revising its 1968 survey in 1973, the Educational Testing Service prepared a profile of state testing programs reported in the 1972-73 school year (ERIC Clearinghouse on Tests, Measurement and Evaluation, 1973). The survey included fifty states, the District of Columbia, Puerto Rico, and the Virgin Islands. Forty-two statewide testing programs were found to exist in thirty-three states, and additional programs were being planned. Some states had multiple testing programs. For example, New York reported seven; Florida reported two; and Puerto Rico reported two. The most common goal of these programs was to evaluate instruction. Testing was administered in grades eight through twelve in sixty-four programs and in grades one through seven in fifty-nine programs.

The most consistent constellation of subject areas tested consisted of reading, mathematics, and language skills. This core plus one other area formed a pattern in at least one-third of the states. Language skills were tested in twenty-seven programs. No delineation of speaking and listening per se was found, but states did use different terms to refer to language skills:

English	13 states
Language arts	5 states
Language	2 states
Vocabulary	1 state
English expression	1 state
English skills	1 state

Testing instruments used were Differential Aptitude Tests, Comprehensive Test of Basic Skills, Iowa Tests of Educational Development, SRA Achievement Series, California Achievement Tests, School and College Ability Tests. No single test occupied a dominant position; but given our instrument review, we can see assessment of speaking and listening would not figure heavily in the program. Related skills such as vocabulary usage and verbal reasoning were measured by a number of tests, but they were assessed more in the context of writing.

In 1973 the Educational Testing Service also revised its earlier survey of statewide assessment programs reported in the 1972-73 school year (Center for Statewide Educational Assessment, 1973). According to this survey, forty-three states had established educational goals, while seven other states were in the process of doing so. Only four -- Missouri, Montana, Oklahoma, and Utah -- listed communication skills as an area to be assessed. Instrumentation for assessing these and other skills was not yet designed in Missouri and Montana. Instrumentation in Oklahoma was described as including three different survey booklets specifically prepared for assessment. Instrumentation in Utah was described to include local testing programs.

From these ETS surveys, Grommon (1976b) cited five positive aspects and eight problems of statewide programs of testing and assessment. These are paraphrased as follows:

Positive Aspects

1. States formulate educational goals acceptable to a representative segment of concerned people in the state.
2. Most states no longer consider schools and the teacher responsible for teaching solely the three R's.
3. Citizen advisory groups, teachers and students were/are used to develop educational goals.
4. An increasing number of states are either already using standardized tests "tailor-made" to fit identifiable circumstances or are in the process of creating such tests.
5. Many states have replaced conventional standardized norm-referenced measures with criterion-referenced tests; some states are in the process of doing so; and others are using criterion referenced tests to supplement norm-referenced ones.

Serious Problems

1. State programs can be/are being construed as threats to the local control of schools.
2. The use of subject-matter tests in isolation to evaluate instruction, programs, and educational planning are likely to be resisted as misapplication of test results.
3. Statewide testing programs develop negative attitudes among teachers towards all testing and particularly towards standardized tests. This subverts the purposes of testing.
4. There is basic incompatibility between the use of a single specific standardized test and evaluation of students' experiences in an array of courses.

5. There is a significant discrepancy between a teacher's concept of a subject and a concept that underlies a standardized test.
6. Linguistic concepts represented by items of some standardized tests constitute a crucial problem. Cultural bias contaminates linguistic items in a test or indeed the whole test.
7. The use of standardized tests encourages teachers to reduce teaching to the simplistic level of quantifiable behavioral objectives. Teachers protest this.
8. The practice of comparing standardized test results to so-called "national" norms and then projecting comparisons into judgments about the quality of teaching ignores inescapable factors such as the local learning environment.

Grommon's positive aspects refer more to assessment programs, whereas his list of problems is more of a commentary on testing programs.

The 1976 report of the National Association of Secondary School Principals (Clark and Thomson, 1976) indicates that the use of tests to verify competencies required for graduation is a relatively new but growing movement in the United States. Only Arizona required verified competencies for the class of 1976 (aside from the general equivalence test developed by California). New York, of course, has administered the Regent's Examinations and has awarded Regent's diplomas for some time.

Clark and Thomson indicate state activity in the minimal competency movement included the following as of 1976:

California	- In process of seeking legislation to require every student to achieve a certain level of competency together with a marketable skill.
Oregon	- In 1973 established six goals for public schools. This lead to the identification of twenty areas of personal, social and career development required for the diploma as of 1978. Oregon includes basic skills under "Personal Development." Basic skills are: reading, writing, computing, listening, speaking, and analyzing.
Virginia	- Effective in 1976 requires minimal objectives and testing in functional literacy, computational skills, U.S. history and culture, and either ability to enter post-secondary school or attainment of a job skills for immediate employment.
Florida	- In 1975 proposed a special diploma certifying minimum competency in basic skills.
New York	- Measures basic competencies in reading, mathematics, civics and citizenship, practical sciences and health, and writing skills. As of 1979, reading and mathematics tests are required for the diploma.

Louisiana - Proposed statewide competency requirements in reading and mathematics.
Maryland - Is considering various forms of proficiency and exit exams.

From this review, the NAASP determined that four states, California, Colorado, Florida, and Virginia enacted legislation regarding minimal competency testing. Legislation was pending in eight states. And state board rulings had occurred in eight states.

Pipho (1978) reported that in 1978 minimum competency testing looked different than it did in 1975, 1976, or 1977. He states:

> What began as a startling idea in California, Florida, Oregon, and a handful of other states in 1975 and 1976 has now arrived in some form in each state. As of March 15, 1978, thirty-three states had taken some type of action to mandate the setting of minimum competency standards for elementary and secondary students. All the remaining states either have legislation pending or legislative or state board studies under way. (p. 585)

State-by-State Review

Pipho reported his most recent study of state activity on minimal competency tests in January of 1979. He indicated thirty-seven states were assessing given skill areas. Table 10, based on Pipho's 1979 survey, shows that, other than Massachusetts, the following states have developed minimal competencies in speaking and/or listening: Vermont, Virginia, Oregon, and Utah.

Vermont (Vermont Department of Education, 1977; 1978a) includes seven speaking and three listening competencies as follows:

Speaking
- giving geographical directions
- organization, sequence, clarity and accuracy in explaining a process, making a report, or expressing an opinion
- answer a business telephone and take a message
- get information using a telephone
- introduce self and others
- job interviewing
- participate in and listen to others in informal group discussions

Listening
- follow three-step directions
- retell important events in a story sequentially
- summarize essential details of an oral presentation

Assessment of these competencies is accomplished through three means: reliance on teacher judgment, teacher maintenance of reports and records, and validation by a state review team. Let us examine these means further.

Table 10. Skill Areas Assessed in State Minimal Competency Testing Programs

State	SKILL AREAS								
	Reading	Writing	Math, arithmetic, computation	Language arts, grammar, spelling	Speaking	Listening	Social studies, consumerism, economics, democratic governance	Science	Life skills, survival skills
AL[1]									
AZ	x	x	x						
CA	x	x	x				x		
CO[2]									
CN[3]	x		x	x					
DE	x	x	x						
FL[4]									
GA	x		x				x	x	
ID	x	x	x	x					
IL[5]									
IN	x								
KA	x		x						
KY	x	x	x	x					
LA[6]			x						
ME[7]									
MD	x								
MA	x	x	x		x	x			
MI	x		x						
MO	x		x				x		
NE	x	x	x						
NV	x	x	x						
NH	x	x	x						
NJ	x		x						
NM[8]									
NY[9]	x	x	x						
NC[10]									
OR[11]	x	x	x						
OK	x		x						
RI									
SC	x	x	x						
TN	x		x	x					
UT[12]	x	x	x		x	x			x
VT	x	x	x		x	x			
VA	x	x	x		x	x			
WA	x		x	x					
WY	x	x	x				x		

---

1. AL lists "Basic skills state test."
2. CO lists "Local district option."
3. CN also lists "General Education Development (GED) Test."

Table 10. Skill Areas Assessed in State Minimal Competency Testing Programs (continued)

4. FL lists "Basic skill areas" and "functional literacy."
5. IL lists "Local districts with assistance from the State Board of Education."
6. LA also lists "communication" skill.
7. ME does not list areas to be tested. These areas are to be determined by a statewide committee and the Commissioner of Education.
8. NM indicates "local option" on elementary level; proficiency battery based on adult performance level (APL) and a writing sample.
9. NY indicates evaluation by Pupil Evaluation Program (PEP) or by passing Regent's Examination in English and mathematics. The Regent's Comprehensive Examination in English includes a listening test.
10. NC does not list specific areas; these are to be determined by the Competency Test Commission.
11. OR also lists local option for personal development (which includes speaking and listening), social responsibility, and career development.
12. UT also lists problem solving.

In Vermont teachers determine when and how a student demonstrates achievement of a competency during regular classroom activities. Teachers assess achievement of competencies continuously, but they are supported by state developed assessment guidelines. By accommodating the local needs of pupils and teaching styles of teachers, Vermont recognizes that standardized data are sacrificed, but the Department of Education feels that pupils learn best when their teachers are allowed maximum flexibility and professional judgment. So their data cannot be 100 percent comparable.

The state guidelines for assessment of speaking and listening allow for more subjectivity than in other language arts fields. These guidelines include (a) suggested criteria of a demonstrably functional level of performance for each competency, and (b) sample materials for use in preparing evaluation items. Two examples of suggested criteria are as follows:

Speaking - Explaining a process, making a report, or expressing an opinion.

The pupil may use notes but should not read his/her presentation for this competency. The presentation should be of sufficient length to give the pupil time to develop his/her topic (at least two or three minutes). Organization of the material as well as clarity of presentation should be criteria for evaluation. The presentation may be given to the entire class or to a small group.

Listening- Summarizing factual material presented orally.

The pupil will listen to a short presentation in which several important facts are included. He/she may take notes during the presentation but may not ask for repetition. The pupil will summarize the material in either written or oral form, including at least 80% of the major facts. The material used should not be complicated by complex statistics or numbers and should have an easy-to-follow organization pattern.

The sample materials are primarily an enumeration of situations for assessing each competency. For example, with the speaking competency "giving directions," listed situations are:

A. Direct an out-of-state tourist from the school to a motel.
B. Tell the ambulance how to reach the school or home.
C. Direct a parent to the school office or to a classroom.
D. Direct a person to the nearest entrance to the Interstate.
E. Direct a person to a store, the police station, or other local building.

Teachers are also expected to maintain records and reports as evidence that competencies have been achieved. Using a record that lists the competencies, the teacher records the month and year when a pupil attains mastery of a competency and other information required at the local level, e.g., introductory work, maintenance of mastery, review required to regain mastery. Instructions to the teacher state: "The record should show mastery several times, over more than one year, to indicate maintenance of mastery." Near the end of the year, the teacher completes a tally of the number of pupils she/he has by year of birth and the number that have mastered each competency to date. Supervisors combine these data and submit results to superintendents. These data are then compiled by the Department of Education.

The data reported by the teachers and various superintendencies were validated by the Department of Education through a procedure that involved two studies: a review study and a comparability study. The review study compared findings of a state review team, on a sample of students, with findings of local assessment. For all the competencies reviewed, i.e., reading, writing, speaking, listening, and mathematics, the percent of agreement of local and state review team assessments was 69%; for speaking competencies, the percent of agreement was 66%; for listening competencies, it was 62%. The comparability study judged the extent to which the approaches used in assessments in various school districts were comparable. From both studies, the review team found indications that data at the local level were being used constructively to determine the strengths and weaknesses of individual pupils and to provide opportunities for the remediation of deficiencies.

Virginia's Basic Learning Skills Program is designed for grades K-6 (Virginia Division of Elementary Education, 1977). It includes two major listening and two major speaking skills, each of which are further defined by more specific minimum objectives. Each minimum objective is keyed to a grade range, e.g., K-2, 1-3, 4-6. The four major objectives are:

Listening: 1. The student will comprehend main ideas and specific details of an oral communication.
2. The student will comprehend that the meaning of an oral communication is influenced by many factors.

Speaking: 1. The student will use correct and appropriate language.
2. The student will speak effectively in a variety of situations.

There are eight specific minimum objectives for listening goal #1; eleven for listening goal #2; six for speaking goal #1; and 16 for speaking goal #2; forty-one specific minimum objectives in all.

According to Mr. Claude Sandy (1979) of the Virginia State

Department of Education, there has been no development of a formal test or assessment procedure for speaking and listening to date.

Oregon's minimum competencies include demonstration of ability to read, write, speak and listen. (Oregon Department of Education, 1978a and b). School districts must include language arts/English in their K-12 curriculum. Suggested speaking and listening goals are:

- -Speaking effectively when addressing different purposes and different audiences;
- -Using language effectively in interacting with individuals and groups;
- -Using appropriate conventions in communicating ideas clearly and concisely;
- -Listening purposefully in acquiring, interpreting, and evaluating information;
- -Using nonverbal communication to express ideas, attitudes, feelings;
- -Describing a variety of ways in which ideas are expressed.

Assessment procedures are determined at the local level. School districts must assess group performances only in reading, writing, and computing.

Utah requires two kinds of minimal competencies: basic skills and life skills (Utah State Board of Education, 1977, 1978). Basic skills competencies are reading, writing, speaking, listening, computation, and problem solving. Life competencies are in four areas: consumer of goods and services; career; health and safety; and democratic governance. Both sets of competencies are interrelated. For example, there are four speaking skill competencies. Each one is related to a life competency as follows:

Consumer of goods and services
Negotiates business transactions, expresses needs and conducts other consumer related oral communication with sufficient clarity, volume, and purpose.

Career
Conducts career related oral communication, e.g., responding to questions in a job interview with sufficient clarity, volume, and purpose.

Health and safety
Explains symptoms to doctor, requests safety information, and conducts other health and safety related oral communication with sufficient clarity, volume, and purpose.

Democratic governance
Expresses views, questions officials, and conducts other citizenship related oral communication with sufficient clarity, volume, and purpose.

The listening skills in each life competency area are:

Consumer of goods and services
Listens and responds appropriately to commercials, instructions, and other oral communication necessary as a consumer of goods and services.

Career
Listens and responds appropriately to questions and other oral communications related to career.

Health and safety
Listens and responds appropriately to doctors' instructions, safety warnings, and other oral communications related to health and safety.

Democratic governance
Listens and responds appropriately to statements by elected officials, participants in a discussion, and other citizenship related oral communication.

Preliminary assessment plans call for testing of basic academic skills by commercially produced achievement tests at grades 3, 5, 8 and 11. Competency in the twenty-four areas required for graduation is to be assessed by prescribed methods as a regular part of classroom activities in grades nine through twelve. Assessment methods are detailed in "Functional Competencies for High School Graduation: Teachers' Handbook 1978/1979." Materials in this book are in preliminary form; they are yet to be field tested with appropriate reliability and validity checks.

Utah's prescription for evaluating each competency includes proficiency level, curriculum location, and five performance indicators. Here is one example for speaking related to the life competency area "consumer of goods and services":

Proficiency Level:	Negotiates business transactions, expresses needs and conducts other consumer related oral communication with sufficient clarity, volume, etc.
Curriculum location:	The curriculum areas which provide instruction in the skills, knowledge, attitudes, etc., related to this proficiency, and the class where the competency will be tested are to be identified by the District.
Performance indicator:	Given information from a label concerning a specific product, the student will orally describe the product and its uses. Communication must be audible, pertinent to the subject, and without the use of outside resources. (Four additional performance indicators are listed for this proficiency level.)

While Pipho (1979) reports only these four states -- Vermont, Virginia, Oregon and Utah -- require minimal skills in speaking and listening, we found related activities in other states.

The Michigan Department of Education (1978) proposes twenty minimal competencies in five areas of oral communication: basic principles of conversation, group discussion, creative dramatics, public speaking, and critical listening. The state is currently weighing methods of assessing speaking, but a listening test has already been developed for grades four, seven, and ten. Three forms of this instrument will be field-tested in the 1979-80 school year. Our review of the instrument was reported in Chapter II.

Alabama (Lieb-Brilhart, 1979) is currently field testing instruments that assess speaking and listening skills in grades four, eight, and twelve. The specific listening skills vary with grade level but are generally organized around: discrimination, recall, comprehension, and evaluation. Speaking skills stress language, voice and diction, and message organization.

Both sets of skills are assessed in a complementary way in fourth grade. Using two tape recorders, one called the "test tape" and one the "student tape," the student listens to stimulus materials from the test tape and responds on the student tape. The stimulus materials begin with brief items that test listening; then they require the student to talk beyond the materials heard on the test tape. Each student's responses -- to the listening items and to the instructions for speaking -- are scored by the teacher using the student tape. Because this procedure is time consuming, test administrators randomly sample students in a class.

At the eighth and twelfth grade levels, means of assessment diverge. Eighth graders listen to a test tape containing a story. They are then given a list of questions and an answer sheet to mark correct responses. They listen to and read each question and mark their answer on the sheet provided. The test tape then instructs them to tell a story based on what they have heard or to answer a series of questions, some of which call for extended responses. Their oral responses are recorded on the student tape and scored.

Twelfth graders take a paper-pencil listening test by responding to questions read aloud by the teacher. For speaking, they participate in a person-to-person interview. Thus, these two test situations are not as interrelated as at lower grade levels.

The California High School Speech Association and the California Speech Association produced a framework to be developed in Grades K-14 (Brosbe, 1973). It contains numerous speaking and listening objectives that are related to cognitive, affective and psychomotor domains.

Pennsylvania's Comprehensive Reading/Communication Arts Plan (PCRP) is intended to advance competence in listening and oral expression as well as in reading and writing (Botel, 1977). The primary means of assessing oral communication is through informal checks at the beginning, middle, and end of the year. By using audio or video tapes at these times, the teacher is encouraged to listen for the quantity, quality and broadness of children's participation.

New York assesses speaking and listening through the Regent's Comprehensive Examination in English and through its Statewide Achievement Examination in English (Jandt, 1979). The Regent's examination tests listening skills by having the student respond to ten questions in a test booklet after two oral readings of a passage by the teacher. Students complete their answers during the second oral reading. The items test for listening to determine the purpose; listening for the main idea; listening for organizational patterns; listening for transitions; listening for specific words; and listening to draw inferences. Different forms of the test are provided.

The Statewide Achievement exam tests listening in the same way as the Regent's exam. It also tests speaking by having the student prepare and present an oral response to an assigned situation or problem. The student selects a card with three problems or situations printed on it. After two minutes of choosing one item planning a response, and jotting down notes (key words and phrases) on a 3" by 5" card, the student makes a two minute presentation.

The examiner rates the student's talk as a whole on a ten point scale as follows:

Nonperformance	Weak	Adequate	Strong
0	1-6	6½ - 7½	8-10

Criteria for rating are organized around content, organization, and voice and articulation. The examiner's manual contains sample questions, responses, and ratings as a guide. It also contains a master list of the forty-five stimulus problems or situations. Instructions to the examiner indicate pupil responses may be taped for future reference, but taping is not to interfere with pupil responses.

The Instructional Objectives Exchange (1972) of Los Angeles, California, presents thirty-one listening objectives and fifty speaking objectives. Each objective is accompanied by the description of a situation, criteria for evaluating student responses, a sample test item, and a possible answer. This exchange is intended to provide resource materials for classroom teachers throughout the country. It is not primarily concerned with testing for minimal competencies per se.

Local Jurisdictions

Concern for minimal competencies is also reflected at the local school district level. The Parkrose School District of Portland, Oregon, requires students to demonstrate two major listening competencies and four major speaking competencies (Mershon, 1979). Each competency is defined further by several specific "indicators." The school district provides courses in speaking and listening at both the junior and senior high school levels. Successful completion of the courses constitutes sufficient evidence that the minimum competencies have been met. Thus, speech teachers certify that students have met the competencies through the regular instructional program.

The Lincoln, Nebraska schools employ the speaking/listening competencies developed by the Speech Communication Association (Beall, 1979; see also Bassett, Whittington, and Stanton-Spicer, 1978). Each competency is assessed informally in the elementary grades and more formally in the junior and senior high school years. "Formal" assessment includes evidence that the student has achieved more than fifty per cent of a set of specific indicators for each of the ten major competencies. Also students who have successfully completed speech, debate, or drama courses are judged as having achieved the given competencies.

The Westside Community Schools of Omaha, Nebraska, require oral communication minimal competencies (Findley, 1979). In the second semester of their sophomore year, students are tested on their ability to present ideas orally. The test situation may be a formal or informal talk. Seven competencies are assessed. All emphasize speaking more than listening. The test is administered by the student's homeroom advisor. A dichotomous rating scale (yes-no) is employed. Students must receive 100% "yes" responses before the minimum competency is met. James Findley, Assistant Principal of Westside High School, indicates that the lack of rater reliability is the "biggest weakness" with the oral communication competencies.

Two school districts, Aurora, Colorado (Aurora Public Schools, n.d.) and Bellevue, Washington (Bellevue Public Schools, n.d.) prescribe speaking and listening competencies, but neither one indicates how the competencies are assessed.

The Gary, Indiana school district assesses students' oral proficiency in seven areas: articulation, pronunciation, verbal utterances, rate, word usage, voice qualities, and volume (Gary Community School Corporation, 1977-78; Hall, 1979). Students are assessed in the sophomore level "General Speech" course. While this course includes listening, listening is not assessed as a part of oral proficiency.

In Gary, oral proficiency was tested originally in one of four situations selected by the student: an interview with a personal interest, educational, or biographical theme; a topic format stressing an impromptu response; a question format; and a prepared speech format. Currently students are evaluated only in an interview situation (Hall, 1979). The examining team of two speech teachers randomly selects three to five questions from a master list of thirty-two. Student responses are rated holistically for each proficiency on a four point scale:

1 = severely deficient
2 = deficient to moderately deficient
3 = average proficiency
4 = moderate to high proficiency

Using this scale, ratings of one and two are failing. Ratings of three and four are passing. Audio tapes of student responses are kept confidential and are used only upon student/parent objection to the score. Teachers do not rate students in their own schools.

Summary of Practices

The practices described in the previous section are summarized below in two categories: prevalence of competency and assessment programs, and nature of assessment practices.

Prevalence

1. At least thirty-seven states are currently implementing minimal competencies.
2. At least four states have adopted competencies in speaking/listening. Five other states report activities related to the development of competencies in speaking/listening.
3. Assessment of these competencies is vested with local school districts in all of the states that have adopted competencies.
4. Two of the four states (Vermont and Utah) provide definite guidelines for assessment of basic competencies. In Utah efforts to assure assessment reliability and validity are being planned. In Vermont, validation efforts on a sampling of students indicates (a) an agreement range of 62%-68% between teacher judgment and judgment by state review teams, and (b) variations in assessment materials and methods.
5. One state, New York, tests students in speaking/listening in the Regent's and Statewide examinations. The listening exam appears to be a standardized norm-referenced test. The speaking exam appears to be one requiring a performance that is graded holistically by a single examiner. Two additional states are field testing instruments.

Nature of Assessment Practices

6. Informal assessment is suggested for the elementary grades. Formal assessment occurs between grades seven and ten.
7. Assessment of speaking and listening is generally viewed as more subjective than assessment of basic skills in reading and mathematics, but efforts are under way to increase objectivity.
8. At the local district level certification of competency achievement occurs through successful completion of given speech courses (Lincoln, Parkrose) and through specially scheduled testing situations (Westside and Gary).

9. While both speaking and listening are included in minimal competency programs, more attention has been given to assessment of speaking than to assessment of listening.

10. Assessment situations include: (a) specially developed norm-referenced tests and prescribed proficiency situations for listening; (b) interviews, informal impromptu talks, prepared extemporaneous speeches, and responses to simulated and real problems for speaking. On a continuum from highly structured situations (prepared formal speeches) to unobtrusive situations (assignments that occur in the normal routine of the classroom), practices vary; but semi-structured situations (interviews, small group talks, impromptu responses in a pre-arranged assessment session) are most prevalent.

11. Stimulus probes for listening include taped and/or written questions and oral instructions read aloud by the teacher. For speaking, probes range from open-ended assignments (e.g., communicate orally, so that you can be easily understood, an idea of your choice to a group of your peers and at least one teacher) through predetermined general questions about personal, biographical, educational, vocational, social and civic interests and beliefs (e.g., tell us about some of your ambitions for the future; explain some changes you would make if you were principal of this school) to specific problems that require a solution (e.g., what information is needed by a person who is thinking about buying a car, boat, motorcycle or snowmobile; direct an out-of-state tourist from the school to a motel; given a simulated accident with personal injury, describe basic first aid steps to be taken).

12. Examiners and raters are usually teachers. Frequently these are speech teachers, but teachers of other subjects are used. Student or peer evaluations are not utilized.

13. Usually only one speaking sample is used to determine competency achievement. This practice can be attributed partly to logistical problems of testing. But it can also be attributed to the specificity of competencies being assessed. When competencies focus on elements common to different communication purposes and life situations, one sample is considered sufficient for valid and reliable assessment. However, when competencies are related to different communication purposes, e.g., describing, giving directions, explaining, or to different life situations, e.g., being a consumer, finding and maintaining a job, sampling is more of a complex problem.

14. Usually the judgment of competency achievement rests with one rater. Multiple raters are used in rare instances.

15. While achievement of competencies is generally rated holistically, different methods of holistic evaluation are used. Departing from the ETS definition of holistic evaluation (Fowles, 1979) and from Lloyd-Jones's (1977) differentiation between atomistic and holistic tests, Cooper (1977) defines seven types of holistic evaluation: essay scale, analytic scale, dichotomous scale, feature analysis, primary trait scoring, general impression marking, and "center of gravity" response. Practices in states and local districts include use of analytic scales, dichotomous scales, and general impression marking.

16. Criterion-referenced measures are employed to evaluate speaking behaviors, whereas both criterion-referenced and norm-referenced measures are employed for listening.

17. The interpretation of rating results varies with the type of evaluation used. In the use of dichotomous scales, 100% achievement is required for passing. In general impression marking, students must achieve at least 6.5 on a 10 point scale. With an analytic scale, students must score a total of 24 points (of a possible 28) and not below a 3 (on a 4 point scale) in more than 2 categories (of a possible 7) or below a 2 in more than two categories.

Comparison of Practices with Criteria[1]

The practices reported in other states and jurisdictions generally compare favorably with the content specifications of our criteria. Assessment of speaking/listening performance (#1) is emphasized. Assessment focuses on given competencies (#4) in familiar situations (#5) that allow for a range of responses (#7). While optional settings are not always available to students, there is some indication that the student may choose to structure the situation (#6) to be relatively formal or informal. It is clear that speaking performance is not greatly mediated by the ability to read and write (#2).

Some potential weaknesses do appear. First, assessment of listening is mediated by the ability to read and comprehend written questions (#2). Second, the extent of cultural bias is not clear. We saw little evidence of stereotypic content in instruments, but the expectation that students will demonstrate proficiency in standard English usage is common. This expectation can be construed as a potential source of bias. Beyond this, the extent of bias is dependent on the compatibility between the rater's and student's backgrounds and the extent to which the rater separates the content of student responses from the competencies to be demonstrated.

Practices in other states and jurisdictions also compare favorably with a number of the technical specifications in our criteria. Assessment is keyed to predetermined competencies (#10) in efficient and practical ways (#13) that involve equipment no more complicated or expensive than audio-tape recorders (#14). Also assessment is keyed to the target population (#15) who has some influence (#12) over the formality of the testing situation.

However, three problems do exist. First, there may be a problem with concurrent validity. Second, while there are efforts to standardize test administration, the result of these efforts is not clear. Third, inter-rater reliability is the most glaring weakness. Only Gary, Indiana has attempted to meet this problem "head-on." The most prevalent practice seems to be to admit the subjectivity of rater responses and to do little about it.

A related problem pertains to the number of speaking/listening samples that are used to ascertain competency. One speaking sample is common practice. Given the "purpose free" nature of competencies in other locales, this single sample may be sufficient. However, the Massachusetts competencies include communicating for different purposes. This may pose sampling problems if each competency must be achieved.

---

[1] In this section, Criteria are noted by number within parentheses.

Parting Comment

While statewide practices in assessing speaking/listening skills appear to be in the infant (or perhaps embryonic) stage in the United States, we note considerable experience in England with examinations of oral expression. Wilkinson (1965) points out that the English Speaking Board as well as other boards have conducted exams in England for years.

By classifying speech testing situations on a continuum ranging from reciprocal speech, i.e., conversation, to formal speech, i.e., the public speech, Wilkinson describes possible test situations such as the interview, group discussion, seminar, lecturette, recitation, dramatic activities, debate, and reading aloud. He further proposes marking schemes for voice, content, fluency, reciprocity, and interpretation. He then reviews over a dozen test batteries, all designed to assess speaking ability. Some criteria for testing suggested by Wilkinson are as follows:

1. A test of oral expression should include a test of ability to use speech in a fairly normal reciprocal situation. Certain formal speaking elements are appropriate, but extreme formality is not.
2. The classroom situation is a good one to keep in mind.
3. While there are doubts about the reliability of continuous assessment, it does have much to commend it, for pupils are considered under unstressed conditions.
4. An over-elaborate examination will defeat its own ends; but speech examinations are more portable than written ones, and the time factor is not insuperable.[1]

While severe time restrictions imposed on this study prevented further investigation of testing practices in England, it is apparent that a separate study of practices in that country might provide additional directions for assessment in Massachusetts and other states.

---

[1] Wilkinson et al. also developed the "Learning Through Listening" instrument we reviewed in Chapter II.

CHAPTER IV

RECOMMENDATIONS FOR THE DEVELOPMENT OF INSTRUMENTS AND PROCEDURES

In Chapter II we reported that no instruments or procedures already developed are suitable for adoption in Massachusetts. In Chapters II and III we also indicated some promising instruments and practices can be identified. Therefore, we recommend the development of new instruments that incorporate these features. Since the Basic Skills Improvement Policy calls for assessing skills not later than the beginning of ninth grade, the kinds of instruments we are recommending should be designed for the eighth grade level.

Our recommendations call for the development of two kinds of instruments: a rating scale to assess speaking skills, and a test to assess listening skills. We organize our recommendations around these different instruments.

Assessing Speaking Skills

1. Assessment of students should occur in naturalistic rather than contrived situations.

a. Contrived situations are those that isolate assessment from regular classroom activities. These situations employ what we referred to in Chapter II as "obtrusive" instruments. One example of assessment in a contrived situation is found in New York (see Chapter IV). While a formal test of speaking may have been deemed appropriate in that state, circumstances in Massachusetts, including the criteria described herein, warrant considering other approaches.

Naturalistic situations integrate assessment in school and classroom activities that are familiar to the student. The practice in Vermont approximates this type of situation most closely.

Some examples of naturalistic situations that correspond with our competencies and that can be monitored in the classroom ask the student to do the following:

Ask for and give straightforward information:

1. Tell how you would direct a visitor to the school office (or some other site).
2. Tell how you would explain to a new student a school policy regarding tardiness, dress code or some other regulation.
3. Tell what you would ask to find out how to join a record club, how to get to the swimming pool, etc.
4. Tell how you would explain a homework assignment to a student who was absent.

Describe objects, events, experiences:
1. Describe to another student how she/he would find a particular book among a number of other books in your locker.
2. Describe to a fellow student an event that you attended recently, e.g., a sports event, a concert, a party.
3. Describe to a new student different places they can go for fun and recreation.

Question other's viewpoint:
1. Question a fellow student on his/her interpretation of a book, movie, TV program she/he read or saw recently.
2. Question someone's position on a school policy, a given value, or tradition.
3. Question an opinion you heard stated on TV or read somewhere.

Using survival words:
1. Tell what you would do if you smelled smoke in the classroom. Would you say anything? If so, what and to whom?
2. Tell what you would do if you saw a friend accidently hit on the head and pass out. Would you say anything? If so, what and to whom?
3. Tell what you would do if you saw a young child walking into a path of a moving car. Would you say anything? If so, what and to whom?

Speak so listener understands purpose:
1. Use the telephone to invite a friend to a party, to obtain a job interview, to order a product, to persuade a friend to go to a school or non-school event.
2. Talk to the teacher for different purposes, i.e., to get information on a homework assignment, to share a joke or anecdote, to question a grade.

Other general activities in which skills can be assessed are participation in small group discussion and oral presentation of different kinds of talks and reports.

b. While we believe the situations described above tap typical communication demands outside as well as inside the school, we emphasize that these situations are not inherently natural in the classroom. If teachers customarily exclude such activities from their teaching methods and practices, students will not see them as familiar. So, for these situations to be natural, it is incumbent on the teacher to include different oral activities among other typical assignments such as writing reports and themes, reading textbooks, and taking quizzes.

2. Assessment of students should involve observations by classroom teachers.

a. Teachers and non-teachers, e.g., other students and parents, are possible observers of student activity. Of these, teachers are the most likely observers. Teachers of different subjects might be recruited to rate students' speaking skills, but we expect that English language arts and speech teachers will be more involved in assessment of communication skills than other teachers.

b. All persons who observe and rate students' speaking skills should be normal hearing persons.

c. All persons who conduct assessment should do so in the course of regular classroom activities. Given a rating scale similar to the one we describe under recommendation number three, the teacher can begin to rate individual students anytime in the course of a semester after she/he is assured that the students see the assessment situations as familiar. In the course of assessment, if the teacher finds a student did not respond in a situation or had no opportunity to respond, it may be necessary to formalize assessment by deliberately scheduling situations, within the classroom, that elicit desired evidence of performance.

d. Teachers should observe and rate students at least twice in a given time period. That is, at a minimum, students should be rated two different times by the same teacher or one time by two different teachers. In the first instance, we have a kind of test-retest reliability check; in the second, an inter-rater reliability check. The given time period will vary with schools. For some it will be a semester; for others, a year.

e. Teacher raters should be encouraged to tape record student responses whenever possible. Tapes would then be available should individual students or their parents desire to hear them. Otherwise, the tapes should be kept confidential.

f. Only teachers who have been trained in using, scoring, and interpreting the rating scale should be responsible for assessment. At the very least, this training should include explanation of the rating scale, practice in using it with a sample population of taped student responses, discussion of student responses, and teacher ratings.

3. The assessment instrument should be a rating scale with features such as the following:

a. Rather than list each skill, the scale should reflect principal domains or factors we believe these skills represent. We see the following domains:

Delivery = skills 2, 3, 4
Language = skills 1, 5, 9
Organization = skills 6, 7, 8
Purpose = skills 10, 11, 12, 13, 14

Research reported in Chapter II supports this feature of rating scales; but the particular domains we cite are based on our judgement, not on a factor analysis of the skills.

b. Each domain should be rated on a four point scale with each interval defined in simple language.

c. The scale should contain items to indicate failure to respond and lack of opportunity to respond.

Combining these recommendations, the scale would contain the following:

Delivery

4 Speaker responds with appropriate volume, rate and distinctness creating an impression of poise
3 Speaker responds with generally appropriate volume, rate and distinctness but with less control
2 Speaker responds with only occasional appropriateness in volume, rate or clarity
1 Speaker responds in a distracting fashion with inappropriate volume, rate and clarity
☐ Did not respond
☐ No opportunity to respond

Language

4 Speaker is consistently intelligible; always uses words and phrases in standard English, creating an impression of fluency
3 Speaker is generally intelligible but shows occasional errors in standard English usage and some hesitation in choosing words and phrases
2 Speaker is moderately intelligible but shows frequent errors in standard usage and much hesitation in choosing words and phrases
1 Speaker is barely intelligible; seldom uses words and phrases in standard English usage and shows much stress in choosing words
☐ Did not respond
☐ No opportunity to respond

Organization

4 Speaker organizes and states message clearly with good supporting detail
3 Speaker organizes and states message in a generally clear manner and has adequate supporting detail
2 Speaker organizes and states message in a generally ambiguous manner and has little supporting detail
1 Speaker demonstrates little or no organization; message lacks order and supporting detail
☐ Did not respond
☐ No opportunity to respond

Purpose

4 Speaker responds directly and clearly to the purpose indicated by the situation. Message contains required information
3 Speaker responds in a generally direct way to the purpose indicated by the situation. Message lacks some required information or contains some irrelevant information

Purpose (continued)

2 Speaker responds in a generally indirect way to the purpose indicated by the situation. Message lacks most of the required content and contains irrelevant or ambiguous content

1 Speaker responds with little or no awareness of or connection to the purpose indicated by the situation

☐ Did not respond

☐ No opportunity to respond

d. Rating instructions should be clarified with accompanying sample responses of students, i.e., written transcripts of students' taped responses.

e. The rating scale should be used to assess students in situations that probe two different speaking purposes such as those described in 1 a. This may influence the test-retest reliability mentioned in 2 d, but it might help ensure validity.

4. While the instrument is being developed, assessment should be verified at the eighth grade level.

a. Rating verification should demonstrate inter-rater reliability. This can be achieved by having an "external" rating of a random sampling of schools and students. The "external" rating should be conducted by a Bureau appointed review team that has been trained in using and administering the instrument.

b. The review team should rate students in classroom situations like those first rated by the teacher.

c. Using the same rating scale as the classroom teacher, the review team should rate students in the sample at least once.

d. Members of the review team might include members of the Bureau and Department of Education staff, teachers from schools other than those being sampled, and teachers from within schools that are being sampled. The latter group should consist of teachers who did not provide the first ratings of the student.

e. Within any one school sampled in the review, the review team would consist of one rater and one rating per student. Hence, with each student who was rated first by his/her classroom teacher and who participated in the review sample, there would be a total of at least three ratings as follows: (1) two different ratings by the same classroom teacher or two different ratings from different teachers, and (b) one rating by one review team member.

5. A validation study should be conducted when the instrument is first being developed. Validation is separate from, but related to, the verification.

a. The purpose of validation is to demonstrate that the speaking competencies are subject to development and education.

b. This purpose can be achieved by rating a younger and an older population than the eighth grade target population.

Assessing Listening Skills

To assess listening, we first explored the feasibility of developing an instrument and procedures that would complement our recommended methods of assessing speaking. We considered having classroom teachers observe and rate students' listening behavior in natural situations. We found this possiblity to be flawed because it does not produce tangible evidence of listening behavior. The most it would yield is tapes of student responses which could be construed as speaking behavior more than listening behavior. We therefore recommend that the Bureau develop a listening test with the following features:

1. Stimulus materials should be taped, spoken messages, not written passages read aloud by the teacher or by a narrator on tape. This recommendation overcomes a major criticism of listening tests, i.e., that their stimulus materials are too similar to the stimulus materials of reading tests. The latter contain passages that are characterized more by literary style than by oral style.

2. Stimulus materials and test items should be keyed to the domains or factors represented in our listening skills. The rationale for this recommendation is similar to that in speaking (see 3 a).

a. We see the following domains:

Listening to detect when the message is enhanced or impeded by the speaker's delivery - skill 2
Comprehending the speaker's purpose - skills 6, 7, 8, 10, 11
Comprehending the speaker's organization - skills 4, 5, 9

Unlike speaking, we do not see a separate language domain. We believe this is embedded within responses to all the stimulus materials. So test items to tap skills such as "recognize words and phrases used by the speaker" and "understand words and ideas" in isolation from other domains strike us as artificial. Such items may test little more than oral vocabulary.

3. The stimulus materials should call for a simple and minimal response. This feature has two implications.

a. There will be two kinds of stimulus materials: messages that suggest a situation -- a newscast, a telephone call, a weather report -- and specific questions about the messages. The latter are the test items. Both the messages and the questions should be produced on tape.

b. The test booklet should contain the test items in writing, but these items should not be obstructed by excessive or difficult reading materials.

The goal here is to minimize the mediation of listening by reading ability. Having test items produced orally as well as in writing allows the student to read along with the narrator's question; but students who have some reading difficulty will still be able to respond to the question since they can hear it.

4. *Stimulus materials* -- *both the messages and the test questions* -- *should be short*. The messages should range from thirty seconds to no more than three minutes. The intent here is to reduce the influence of auditory memory.

5. *The stimulus materials should be interesting*. They should be taken from real-life occurrences that would be meaningful to eighth graders. Three possible sources of stimulus materials are: tapes of TV newscasts, weather reports, commercials, program excerpts, telephone calls; tapes of eight grade students collected during the speaking assessment procedures; tapes of "live" dialogues such as a sales exchange in a store, a conversation between teenagers at an amusement center, a conversation during a "rap" session.

6. *The vocabulary used in the stimulus materials should be controlled*. Vocabulary must be simple to minimize the possibility of testing general verbal knowledge.

7. *Stimulus materials should employ speakers who use standard or "network" English*.

Should the development of two different instruments pose feasibility problems, i.e., problems of time and cost, we remind the Bureau that several listening instruments approach our specifications (see Chapter II). We do not see a perfect match of these instruments with our needs, but the instruments show promise. Contractors should be referred to these measures for test development.

We forward these recommendations with our unanimous endorsement.

REFERENCES

Achievement testing and basic skills. Proceedings of the national conference on achievement testing and basic skills. Washington, D.C.: U. S. Department of Health, Education and Welfare and National Institute of Education, 1979.

Aitken, K.G. Using cloze procedure as an overall language proficiency test. *TESOL Quarterly*, 1977, *11*, 59-67.

Allen, R. R., and Brown, K. L. (Eds.). *Developing communication competence in children*. Skokie, IL.: National Textbook Co., 1976.

Allen, R. R., and Wood, B. S. Beyond reading and writing to communication competence. *Communication Education*, 1978, *27*, 286-292.

Aurora Public Schools. *Language arts guide, middle school, 6-8* (Vol. 2). Unpublished manuscript, Aurora, CO.: Author, n.d.

Bassett, R. E., Whittington, N., and Staton-Spicer, A. The basics in speaking and listening for high school graduates: What should be assessed? *Communication Education*, 1978, *27*, 293-303.

Beall, M. Personal Communication, March 6, 1979.

Becker, S. L. The rating of speeches. *Speech Monographs*, 1962, *29*, 38-44.

Becker, S. L., and Cronkite, G. L. Reliability as a function of utilized scale steps. *The Speech Teacher*, 1965, *14*, 291-293.

Bellevue Public Schools. *Speaking task committee: K-6 language arts project*. Unpublished manuscript, Bellevue, WA.: Author, n.d.

Black, J. K. *Formal and informal means of assessing the communicative competence of kindergarten children*. Unpublished doctoral dissertation, The Ohio State University, 1977. (University Microfilms No. 77-24598).

Blankenship, J. B. *A sense of style*. Belmont, CA.: Dickenson Publishing Co., 1968.

Blankenship, J. B. The contrasts of oral and written language. *Report*, 1974, 4, (3), 1-10.

Bloom, B. S., Hastings, J. T., and Maddaus, G. F. *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill Book Co., 1971.

Bonjean, C., Hill, R., and McLemore, S. D. *Sociological measurement: An inventory of scales and indices*. San Francisco: Chandler Publishing Co., 1967.

Bordie, J. G. *Direct and indirect measures of language proficiency*. Paper presented at the meeting of the Linguistic Association of the Southwest, January, 1973. (ERIC Document Reproduction Service No. ED 111 171).

Botel, M. *A comprehensive reading communication arts plan* (Working ed.). Harrisburg, PA.: Pennsylvania State Department of Education, 1977. (ERIC Document Reproduction Service No. 151 731).

Brooks, K. Some basic considerations in rating scale development: A descriptive bibliography. *Central States Speech Journal*, 1957, *9*, 27-31.

Brosbe, J. (Ed.). *Framework for speech communication in California public schools, grades kindergarten through fourteen*. Unpublished manuscript prepared by the Committee on Criteria Based Instruction of the California High School Speech Association and the California Speech Association, 1973.

Brown, C. T. Response to listening: Complex of activities -- and a unitary skill? *Speech Monographs*, 1967, *34*, 465-466.

Bryan, A. I., and Wilke, W. H. A technique for rating public speeches. *Journal of Consulting Psychology*, 1941, *6*, 80-90.

Buros, O. K. *The eighth mental measurements yearbook*, (Vols. 1 and 2). Highland Park, N.J.: Gryphon Press, 1978.

Burt, M., and Dulay, H. Some guidelines for the assessment of oral language proficiency and dominance. *TESOL Quarterly*, 1978, *12*, 177-192.

Carroll, J. B. Development of native language skills beyond the early years. In C. E. Reed (Ed.)., *The learning of language*. New York: Appleton-Century-Crofts, 1971, 97-156.

Carver, R. P. Analysis of "chunked" test items as measures of reading and listening comprehension. *Journal of Educational Measurement*, 1970, *7*, 141-150.

Cazden, C. *Concentrated vs. contrived encounters: Suggestions for language assessment in early childhood education*. Paper presented at the seminar on Language and Learning in Early Childhood, Leeds Institute of Education, January, 1974.

Center for Statewide Educational Assessment and the ERIC Clearinghouse on Tests, Measurement and Evaluation at Educational Testing Service in collaboration with Educational Commission of the States. *State educational assessment programs, 1973 revision*. Princeton, N.J.: Educational Testing Service, 1973. (ERIC Document Reproduction Service No. ED 080 582).

Chastain, K. D. Testing listening comprehension tests. *TESOL Quarterly*, 1979, *13*, 81-88.

Clark, J. P., and Thomson, S. D. *Competency tests and graduation requirements*. Reston, VA.: National Association of Secondary School Principals, 1976.

Clevenger, T., Jr. Influence of scale complexity on the reliability of ratings. *Speech Monographs*, 1964, *31*, 153-156.

Cohen, D. K., and Haney, W. *Minimums, competency testing and social policy*. Staff circular No. 4. Cambridge, MA.: Huron Institute, 1978.

Conrad, E. E., Rentfrow, R. K., Meredith, K., and Fillerup, J. M. *Use of situational language tasks in an intra-teem and teem versus comparison evaluation*. Unpublished paper, Arizona Center for Educational Research and Development, University of Arizona, 1976.

Cooper, C. R. Holistic evaluation of writing. In C. R. Cooper and L. Odell (Eds.), *Evaluating writing: Describing, measuring, judging*, Urbana, IL.: National Council of Teachers of English, 1977.

Cronen, V. E., and Eckles, G. *Development of a method of evaluating oral performance for the purpose of exemption testing in the General Rhetoric Program: Report to the program director*. Unpublished manuscript, University of Massachusetts, 1975.

DeAvila, E. A., and Duncan, S. E. *Language assessment scales, LAS I, for grades K-5, English/Spanish* (2d. ed.). Corte Madera, CA.: Linguametrics Group, 1977. (ERIC Document Reproduction Service No. ED 154 641).

DeSoto, C. B. The predilection for single orderings. *Journal of Abnormal and Social Psychology*, 1961, *62*, 16-23.

Devine, T. G. Reading and listening: New research findings. *Elementary English*, 1968, *45*, 346-348 .

Devine, T. G. Listening: What do we know after fifty years of research and theorizing? *Journal of Reading*, 1978, *21*, 296-304.

DeVito, J. Comprehension factors in oral and written discourse of skilled communicators. *Speech Monographs*, 1965, *32*, 124-128.

DeVito, J. Psychogrammatical factors in oral and written discourse by skilled communicators. *Speech Monographs*, 1966, *33*, 73-76.

DeVito, J. Levels of abstraction in spoken and written language. *Journal of Communication*, 1967, *17*, 354-361.

Diederich, P. B. *Measuring growth in English*. Urbana, IL.: National Council of Teachers of English, 1974.

Duker, S. *Listening bibliography* (2d. ed.). Metuchen, N.J.: Scarecrow Press, 1968.

Duker, S. *Listening: Readings* (Vol. 2). Metuchen, N.J.: Scarecrow Press, 1971. (a)

Duker, S. *Teaching listening in the elementary school*. Metuchen, N.J.: Scarecrow Press, 1971. (b)

Duncan, S. E., and DeAvila, E. A. *Language assessment scales, level 2, LAS II, for grades 6 and up, English/Spanish* (Examiner's manual). Corte Madera, CA.: Linguametrics Group, 1978. (ERIC Document Reproduction Service No. 154 640).

Educational Resource Information Center Clearinghouse on Tests, Measurement and Evaluation and Office of Field Surveys at Educational Testing Service in collaboration with Conference of Directors of State Testing Programs. *State testing programs, 1973 revision*. Princeton, N.J.: Educational Testing Service, 1973. (ERIC Document Reproduction Service No. ED 087 789).

Ellis, D. S., and Minter, R. How good are debate judges? *Journal of the American Forensics Association*, 1967, *4*, 53-56.

Eslava, R., and Lawson, P. O. A project course in spoken English. *TESOL Quarterly*, 1979, *13*, 65-72.

Fagan, W., Cooper, C., and Jensen, J. *Measures for research and evaluation in the English language arts*. Urbana, IL.: National Council of Teachers of English, 1975.

Fairfield Board of Education. *FLES evaluation: Language skills and pupil attitudes in the Fairfield Connecticut Public Schools*. Unpublished manuscript, Fairfield, CN.: Author, 1968. (ERIC Document Reproduction Service No. ED 023 333).

Findley, C. A. *Dyadic task-oriented communication exercises for teaching and testing in the elementary ESL class*. Paper presented at the meeting of Teachers of English to Speakers of Other Languages, April, 1977. (ERIC Document Reproduction Service No. ED 145 692).

Findley, J. E. Personal communication to Tracy Libros, March 6, 1979.

Fischer, R. A. *Listening comprehension: A cognitive prerequisite for communication*. Paper presented at a meeting of the South Central Modern Language Association, 1978. (ERIC Document Reproduction Service No. ED 154 610).

Follman, J. C., and Anderson, J. A. An investigation of the reliability of five procedures for grading English themes. *Research in the teaching of English*, 1967, *1*, 190-200.

Fotheringham, W. S. A technique for measuring speech effectiveness in public speaking classes. *Speech Monographs*, 1956, *23*, 31-37.

Fowles, M. E. *Manual for scoring the writing sample: Basic skills assessment program*. Reading, MA.: Addison-Wesley Publishing Company, Educational Testing Service, 1979.

Garner, W. R. Rating scales, discriminability, and information transmission. *The Psychological Review*, 1960, *67*, 343-352.

Garrison, K. R., and Garrison, J. P. *Measurement of communication apprehension among children*. Paper presented at the meeting of the International Communication Association, Berlin, June, 1977. (ERIC Document Reproduction Service No. ED 139 055).

Gary Community School Corporation. *Oral proficiency program: Secondary schools, 1977-1978*. Gary, IN.: Author, 1977-1978.

Giffin, K. A study of the criteria employed by tournament debate judges. *Speech Monographs*, 1959, *26*, 69-71.

Granowsky, S. A. *Oral language comprehension of children in grades one through six*. Unpublished doctoral dissertation, Duke University, 1976. (University Microfilms No. 77-01069).

Grommon, A. H. (Ed.). *Reviews of selected published tests in English*. Urbana, IL.: National Council of Teachers of English, 1976. (a)

Grommon, A. H. Statewide accountability programs of testing and assessment. In A. H. Grommon (Ed.), *Reviews of selected published tests in English*. Urbana, IL.: National Council of Teachers of English, 1976. (b)

Grossman, G. B. *A comparison of upper primary language arts goals and testing in selected school districts*. Unpublished doctoral dissertation, University of Houston, 1967. (University Microfilms No. 67-16131).

Guilford, J. P., Christensen, P. R., Taafe, G., and Wilson, R. C. Ratings should be scrutinized. *Educational and Psychological Measurement*, 1962, *22*, 439-447.

Hall, C. Personal communication, May 14, 1979.

Haney, W. *Standards for tests and test use*. Staff circular No. 3. Cambridge, MA.: Huron Institute, 1978. (a)

Haney, W. (Ed.) *Summary of the spring 1978 conference of the National Consortium on Testing*. Cambridge, MA.: Huron Institute, 1978. (b)

Haney, W. *Testing the tests*. Staff circular No. 1. Cambridge, MA.: Huron Institute, 1978. (c)

Haney, W. *A review of publications on testing for parents and the public*. Staff circular No. 5. Cambridge, MA.: Huron Institute, 1979.

Heinberg, P. Relationships of content and delivery to general effectiveness. *Speech Monographs*, 1963, *30*, 105-107.

Henrikson, E. H. The relation among knowing a person, liking a person, and judging him as a speaker. *Speech Monographs*, 1940, *7*, 22-25.

Hisel, W. C. *The informal language inventory: Practical language assessment form for elementary classroom usage*. Unpublished doctoral dissertation, Kansas State University, 1974. (University Microfilms No. 75-14363).

Hufford, R. Toward improved tournament judging. *Journal of the American Forensic Association*, 1965, *2*, 120-125.

Instructional Objectives Exchange. *Language arts: Listening oral expression, and journalism, grades K-12*. Los Angeles, CA.: Author, 1972.

Jandt, F. Personal communication, May 12, 1979.

Johnson, F. C., and Frandsen, K. Administering the Brown-Carlsen listening comprehension test. *Journal of Communication,* 1963, *13*, 38-45.

Johnson, O. G. *Tests and measurements in child development: Handbook II* (Vols. 1 and 2). San Francisco: Jossey-Bass Publishers, 1976.

Johnson, O. G., and Bommarito, J. W. *Tests and measurements in child development: A handbook*. San Francisco: Jossey-Bass Publishers, 1971.

Kelly, C. M. Listening: Complex of activities -- and a unitary skill? In S. Duker (Ed.), *Listening: Readings* (Vol. 2), Metuchen, N.J.: Scarecrow Press, 1971, 213-229.

Knower, F. H. A suggestive study of public speaking rating scale values. *Quarterly Journal of Speech*, 1929, *15*, 30-41.

Knower, F. H. A study of rank-order methods of evaluating performances in speech contests. *Journal of Applied Psychology*, 1942, *24*, 633-644.

Krueger, R. F. The reliability of debate judges. *Gavel*, 1949, *32*, 7-9.

Kruger, A. N. Judging the tournament debate judge. *Gavel*, 1955, *37*, 31-35, 43.

Larson, C. E. Problems in assessing functional communication. Communication Education, 1978, 27, 304-309.

Larson, C., Backlund, P., Redmond, M., and Barbour, A. Assessing functional communication. Falls Church, VA.: Speech Communication Association, 1978.

Lieb-Brilhart, B. Personal communication, April 17, 1979.

Lloyd, Jones, R. Primary trait scoring. In C. R. Cooper and L. Odell (Eds.), Evaluating writing: Describing, measuring, judging. Urbana, IL.: National Council of Teachers of English, 1977.

Loban, W. D. Language development and its evaluation. In A. H. Grommon (Ed.), Reviews of selected published tests in English. Urbana, IL.: National Council of Teachers of English, 1976, 45-51.

Lowe, P., Jr. The oral language proficiency test. Unpublished pamphlet, December, 1976. (ERIC Document Reproduction Service No. ED 153 470).

Lundsteen, S. W. Listening: Its impact at all levels on reading and the other language arts (Rev. ed.). Urbana, IL.: National Council of Teachers of English, 1979.

Madden, J. M., and Bourdon, R. D. Effects of variations in rating scale format on judgment. Journal of Applied Psychology, 1964, 48, 147-151.

Matluck, J. H., and Mace-Matluck, B. The multilingual test development project: Oral language assessment in a multicultural community. Unpublished paper presented at the national conference of TESOL, Los Angeles, March, 1975. (ERIC Document Reproduction Service No. ED 119 498).

McCroskey, J. C., and Camp, L. R. A study of stock issues, judging criteria, and decisions in debate. Southern Speech Journal, 1964, 30, 158-169.

McCroskey, J. C., and Camp, L. R. Judging criteria and judges' bias in debate. Journal of the American Forensic Association, 1966, 3, 59-62.

Mead, N. A. Issues related to a national assessment of speaking and listening skills. Paper presented at the meeting of the Speech Communication Association, Washington, D. C., December, 1977. (ERIC Document Reproduction Service No. 146 656).

Mead, N. A. Issues related to assessing listening ability. Paper presented at the meeting of the American Educational Research Association, Toronto, March, 1978. (a)

Mead, N. A. *Assessing listening ability: Relationships with verbal ability and racial/ethnic bias*. Paper presented at the meeting of the Speech Communication Association, Minneapolis, November, 1978. (b)

Mecham, M. J. Measurement of verbal listening accuracy in children. *Journal of Learning Disabilities*, 1971, *4*, 257-259.

Mershon, C. E. Personal communication, May 16, 1979.

Michigan Department of Education. *Minimal performance objectives for communication skills education in Michigan*. Lansing, MI.: Author, n.d.

Michigan Department of Education. *Communication skills objectives* (Draft). Lansing, MI.: Author, August, 1978.

Miller, G. R. Agreement and grounds for it: Persistent problems in speech rating. *The Speech Teacher*, 1964, *13*, 257-262.

Monroe, A. H., Remmers, H. H., and Lyle, E. V. *Measuring the effectiveness of public speaking in a beginning course*. Bulletin of Purdue University, Vol. 36, No. 1, Lafayette, IN.: Purdue University Press, 1936.

Moslemi, M. H. The grading of creative writing essays. *Research in the teaching of English*, 1975, *9*, 154-161.

Mulaik, S. A. Are personality factors raters' conceptual factors? *Journal of Consulting Psychology*, 1964, *28*, 506-511.

Northwest Regional Educational Laboratory. *Oral language tests for bilingual students: An evaluation of language dominance and proficiency instruments*. Unpublished paper. Portland, OR.: Author, July, 1976. (ERIC Document Reproduction Service No. ED 129 099).

Norvelle, L. Development and application of a method for measuring the effectiveness of instruction in a basic speech course. *Speech Monographs*, 1934, *1*, 41-63.

Olsen, J., and Olsen, R. *Adapting the Ilyn Oral Interview to a mass listening test*. Paper presented at the meeting of Teachers of English to Speakers of Other Languages, New York, March, 1976. (ERIC Document Reproduction Service No. ED 121 111).

Oregon Department of Education. *Elementary-secondary guide for Oregon schools: Part I Minimum standards for public schools*. Salem, OR.: Author, 1978. (a)

Oregon Department of Education. *Elementary-secondary guide for Oregon schools: Part II suggestions*. Salem, OR.: Author, 1978. (b)

Pipho, C. Minimum competency testing in 1978: A look at state standards. Phi Delta Kappan, 1978, 59, 585-588.

Pipho, C. State activity: Minimal competency testing. Unpublished paper, Denver, CO.: Education Commission of the States, 1979.

Price. W. K. The University of Wisconsin speech attainment test. Unpublished doctoral dissertation, University of Wisconsin, 1964.

Purves, A., (Chair), and the Task Force on Measurement and Evaluation in the Study of English. Common sense and testing in English. Urbana, IL.: National Council of Teachers of English, 1975.

Sandy, C. Personal communication, May 14, 1979.

Scott, L. (Ed.) Summary of the fall 1978 conference of the National Consortium on Testing. Cambridge, MA.: Huron Institute, 1978.

Scott, R. L. The objectivity of debate judges. Gavel, 1955, 37, 14-15.

Simon, A., and Boyer, E. G. Mirrors for behavior III: An anthology of observation instruments. Wyncote, PA.: Communication Materials Center, 1974.

Smith, R. G. Development and possibilities of message measurement inventories. Paper presented at the meeting of the Central States Speech Association, Southfield, MI., April, 1977. (ERIC Document Reproduction Service No. ED 136 316).

Snow, B. G., and Perkins, K. The teaching of listening comprehension and communication activities. TESOL Quarterly, 1979, 13, 51-61.

Sousa, R. L. Revision, development and implementation of a bilingual evaluation management system (Vols. 1 and 4). Unpublished manuscripts, April, 1977. (ERIC Document Reproducation Service No. ED 139 282).

Stevens, W. E. A rating scale for public speakers. Quarterly Journal of Speech, 1928, 14, 223-232.

Sticht, T. G. The basic skills: A developmental frame of reference,(Working draft). Unpublished manuscript. Washington, D. C.: National Institute of Education, 1979.

Sticht, T. G., Beck, L. J., Hauke, R. N., Kleiman, G. M., and James, J. H. Auding and reading. (HumRRO Tech. Rep. 74-11). Alexandria, VA.: Human Resources Research Organization, 1974.

Thompson, W. I. Is there a yardstick for measuring speaking skill? Quarterly Journal of Speech, 1943, 29, 67-79.

Utah State Board of Education. *State high school graduation requirements and state program of studies in secondary schools*. Salt Lake City, UT.: Author, 1977.

Vermont Department of Education. *Basic competencies: A manual of information and guidelines for teachers and administrators*. Montpelier, VT.: Author, 1977.

Vermont Department of Education. *Language arts curriculum guide*. Montpelier, VT.: Author, 1978. (a)

Vermont Department of Education. *Vermont's 1977-1978 basic competency program: A report on the first school year of implementation*. Montpelier, VT.: Author, 1978. (b)

Vermont Department of Education. *Basic competencies: Teacher's guide for the basic competencies in reasoning*. Montpelier, VT.: Author, 1979.

Virginia Division of Elementary Education. *Basic learning skills, grades K-6*. Richmond, VA.: Department of Education, 1977.

Waldkoetter, R. O., Urrey, V. W., and Martinson, W. D. Some responses to rating stimuli by practiced raters. *Perceptual and Motor Skills*, 1965, *20*, 993-996.

Webb, E. J., Campbell, D. T., Schwartz, R. D., and Sechrest, L. *Unobtrusive measures*. Chicago: Rand McNally and Co., 1965.

Wiemann, J. M. Needed research and training in speaking and listening literacy. *Communication Education*, 1978, *27*, 310-315.

Wilkinson, A. (Ed.). *Spoken English* (Educational Review occasional publications number two). Edgbaston, Birmingham, England: University of Birmingham, 1965.

Wilkinson, A. *The foundations of language*. London: Oxford University Press, 1971.

Wilkinson, A., Stratta, L., and Dudley, P. *Learning through listening* (Teachers' manual). Basingstoke, England: Macmillan Education, Ltd., 1976.

Williams, F., Clark, R. A., and Wood, B. S. Studies in the dimensionality of debate evaluation. *Journal of the American Forensic Association*, 1966, *3*, 95-103.

Wood, B. S. (Ed.). *Development of functional communication competencies: Pre-K - Grade 6*. Falls Church, VA.: Speech Communication Association, 1977. (a)

Wood, B. S. (Ed.). *Development of functional communication competencies: Grades 7-12*. Falls Church, VA.: Speech Communication Association, 1977. (b)

Work, W. ERIC Report: Testing - who's being tested, by whom, by what means, for what, and how well? ERIC looks at the assessment of communication instruction. Communication Education, 1977, 26, 338-344.

Yesulaitis, M. P. C. Procedures for developing oral language facility in children pre-k through grade three: A national survey. Unpublished doctoral dissertation, University of North Carolina at Chapel Hill, 1972. (University Microfilms No. 73-16536).